FIG. 1

Nucleic Acid Sequences

A. predicted cDNA sequence of AtFtn2 (SEQ ID NO:1) (synonym: At5g42480; synonym: *ARC6*) gene

Sequence length = 2406 nt
Start codon (ATG) is at position 1-3
Stop codon (TAA) is at position 2404-2406

```
1    ATGGAAGCTC TGAGTCACGT CGGCATTGGT CTCTCCCCAT TCCAATTATG CCGATTACCA
61   CCGGCGACGA CAAAGCTCCG ACGTAGCCAC AACACCTCTA CAACTATCTG CTCCGCCAGC
121  AAATGGGCCG ACCGTCTTCT CTCCGACTTC AATTTCACCT CCGATTCCTC CTCCTCCTCC
181  TTCGCCACCG CCACCACCAC CGCCACTCTC GTCTCTCTGC CACCATCTAT TGATCGTCCC
241  GAACGCCACG TCCCCATCCC CATTGATTTC TACCAGGTAT TAGGAGCTCA AACACATTTC
301  TTAACCGATG GAATCAGAAG AGCATTCGAA GCTAGGGTTT CGAAACCGCC GCAATTCGGT
361  TTCAGCGACG ACGCTTTAAT CAGCCGGAGA CAGATTCTTC AAGCTGCTTG CGAAACTCTG
421  TCTAATCCTC GGTCTAGAAG AGAGTACAAT GAAGGTCTTC TTGATGATGA AGAAGCTACA
481  GTCATCACTG ATGTTCCTTG GGATAAGGTT CCTGGGGCTC TCTGTGTATT GCAAGAAGGT
541  GGTGAGACTG AGATAGTTCT TCGGGTTGGT GAGGCTCTGC TTAAGGAGAG GTTGCCTAAG
601  TCGTTTAAGC AAGATGTGGT TTTAGTTATG GCGCTTGCGT TTCTCGATGT CTCGAGGGAT
661  GCTATGGCAT TGGATCCACC TGATTTTATT ACTGGTTATG AGTTTGTTGA GGAAGCTTTG
721  AAGCTTTTAC AGGAGGAAGG AGCAAGTAGC CTTGCACCGG ATTTACGTGC ACAAATTGAT
781  GAGACTTTGG AAGAGATCAC TCCGCGTTAT GTCTTGGAGC TACTTGGCTT ACCGCTTGGT
841  GATGATTACG CTGCGAAAAG ACTAAATGGT TTAAGCGGTG TGCGGAATAT TTTGTGGTCT
901  GTTGGAGGAG GTGGAGCATC AGCTCTTGTT GGGGGTTTGA CCCGTGAGAA GTTTATGAAT
961  GAGGCGTTTT TACGAATGAC AGCTGCTGAG CAGGTTGATC TTTTTGTAGC TACCCCAAGC
1021 AATATTCCAG CAGAGTCATT TGAAGTTTAC GAAGTTGCAC TTGCTCTTGT GGCTCAAGCT
1081 TTTATTGGTA AGAAGCCACA CCTTTTACAG GATGCTGATA AGCAATTCCA GCAACTTCAG
1141 CAGGCTAAGG TAATGGCTAT GGAGATTCCT GCGATGTTGT ATGATACACG GAATAATTGG
1201 GAGATAGACT TCGGTCTAGA AAGGGGACTC TGTGCACTGC TTATAGGCAA AGTTGATGAA
1261 TGCCGTATGT GGTTGGGCTT AGACAGTGAG GATTCACAAT ATAGGAATCC AGCTATTGTG
1321 GAGTTTGTTT TGGAGAATTC AAATCGTGAT GACAATGATG ATCTCCCTGG ACTATGCAAA
1381 TTGTTGGAAA CCTGGTTGGC AGGGGTTGTC TTTCCTAGGT TCAGAGACAC CAAAGATAAA
1441 AAATTTAAAC TCGGGGACTA CTATGATGAT CCTATGGTTT TGAGTTACTT GGAAAGAGTG
1501 GAGGTAGTTC AGGGTTCTCC TTTAGCTGCT GCTGCAACTA TGGCAAGGAT TGGAGCCGAG
1561 CATGTGAAAG CTAGTGCTAT GCAGGCACTG CAGAAAGTTT TTCCTTCCCG CTATACAGAT
1621 AGAAACTCGG CTGAACCCAA GGATGTGCAA GAGACAGTGT TTAGTGTAGA TCCTGTTGGT
1681 AACAATGTAG GCCGTGATGG TGAGCCTGGT GTCTTTATTG CAGAAGCTGT AAGACCCTCT
1741 GAAAACTTTG AAACTAATGA TTATGCAATT CGAGCTGGGG TCTCAGAGAG TAGCGTTGAT
1801 GAAACTACTG TTGAAATGTC CGTTGCTGAT ATGTTAAAGG AGGCAAGTGT GAAGATCCTA
1861 GCTGCTGGTG TGGCAATTGG ACTGATTTCA CTGTTCAGCC AGAAGTATTT TCTTAAAAGC
1921 AGCTCATCTT TTCAACGCAA GGATATGGTT TCTTCTATGG AATCTGATGT CGCTACCATA
1981 GGGTCAGTCA GAGCTGACGA TTCAGAAGCA CTTCCCAGAA TGGATGCTAG GACTGCAGAG
2041 AATATAGTAT CCAAGTGGCA GAAGATTAAG TCTCTGGCTT TTGGGCCTGA TCACCGCATA
2101 GAAATGTTAC CAGAGGTTTT GGATGGGCGA ATGCTGAAGA TTTGGACTGA CAGAGCAGCT
2161 GAAACTGCGC AGCTTGGGTT GGTTTATGAT TATACACTGT TGAAACTATC TGTTGACAGT
2221 GTGACAGTCT CAGCAGATGG AACCCGTGCT CTGGTGGAAG CAACTCTGGA GGAGTCTGCT
2281 TGTCTATCTG ATTTGGTTCA TCCAGAAAAC AATGCTACTG ATGTCAGAAC CTACACAACA
2341 AGATACGAAG TTTTCTGGTC CAAGTCAGGG TGGAAAATCA CTGAAGGCTC TGTTCTTGCA
2401 TCATAA
```

B. Genomic sequence of AtFtn2 gene (SEQ ID NO:2) synonym: At5g42480; synonym: *ARC6*)

Sequence length = 3667 nt
This sequence contains 480 nt of the 5' and 149 nt of the 3' region
Start codon (ATG) is at position 481-483
Stop codon (TAA) is at position 3516-3518

```
1    TGTTCTGCAT TAAGGAGAAT ACAATTATAA GCAATTTGTC TTGATTTCAA CAAGATTTTG
61   CTTGGCTATA GGATTCATTG GCTCTGTTTG CTTTTACATT TACATGTCAT AATAGTTTCG
121  AATTTTACAC ATTTCAGTTG GATGTTAAGA AAAGAGAGGG AATTGATGGG GTTTTGTGGG
181  TTTAAACTTT AAAGTAGTCA AGAATTAAGT CATTGGTTTA CTGTTGCTCT ATATGTGTAA
241  AATGAAGGCA ACTCCAACGG TTCTTAGGTG GAATAGATTA TTTAGACGAT TTAACATCAT
301  AAAGTCCGTG GCGACTGTAA CATCATAGAT TGTTTTTTAT TTTTTTCAGT AGCTGGTGAT
361  GTTTTTTGAT TTAACTTATA CTACTCAAAA TCAAAATTCC ATAAACCCTA GACGACCAAA
421  CAGTCTCTTC AATATGTAAA ACAGAACAAA GTTTTTGTAG TAGCCTAAAA AGACACTCCC
481  ATGGAAGCTC TGAGTCACGT CGGCATTGGT CTCTCCCCAT TCCAATTATG CCGATTACCA
541  CCGGCGACGA CAAAGCTCCG ACGTAGCCAC AACACCTCTA CAACTATCTG CTCCGCCAGC
601  AAATGGGCCG ACCGTCTTCT CTCCGACTTC AATTTCACCT CCGATTCCTC CTCCTCCTCC
661  TTCGCCACCG CCACCACCAC CGCCACTCTC GTCTCTCTGC CACCATCTAT TGATCGTCCC
721  GAACGCCACG TCCCCATCCC CATTGATTTC TACCAGGTAT TAGGAGCTCA AACACATTTC
781  TTAACCGATG GAATCAGAAG AGCATTCGAA GCTAGGGTTT CGAAACCGCC GCAATTCGGT
841  TTCAGCGACG ACGCTTTAAT CAGCCGGAGA CAGATTCTTC AAGCTGCTTG CGAAACTCTG
901  TCTAATCCTC GGTCTAGAAG AGAGTACAAT GAAGGTCTTC TTGATGATGA AGAAGCTACA
961  GTCATCACTG ATGTTCCTTG GGATAAGGTA ATTTCGATTT CGGAATAATA AAGTTTCTTC
1021 GTTTTAATTT CATGAATTGG ATAAAGGAAG GAACTTTTAT CTAGTGAAGG TTCCTGGGGC
1081 TCTCTGTGTA TTGCAAGAAG GTGGTGAGAC TGAGATAGTT CTTCGGGTTG GTGAGGCTCT
1141 GCTTAAGGAG AGGTTGCCTA AGTCGTTTAA GCAAGATGTG GTTTTAGTTA TGGCGCTTGC
1201 GTTTCTCGAT GTCTCGAGGG ATGCTATGGC ATTGGATCCA CCTGATTTTA TTACTGGTTA
1261 TGAGTTTGTT GAGGAAGCTT TGAAGCTTTT ACAGGTAGTT TGACTTGCTT TGGTAATTTG
1321 ACGAGCGTTG GCTTTATAAG AACTTTCTTG ATTTGATACT TTGTTATTGA GTCTTGTGTA
1381 GGAGGAAGGA GCAAGTAGCC TTGCACCGGA TTTACGTGCA CAAATTGATG AGACTTTGGA
1441 AGAGATCACT CCGCGTTATG TCTTGGAGCT ACTTGGCTTA CCGCTTGGTG ATGATTACGC
1501 TGCGAAAAGA CTAAATGGTT TAAGCGGTGT GCGGAATATT TTGTGGTCTG TTGGAGGAGG
1561 TGGAGCATCA GCTCTTGTTG GGGGTTTGAC CCGTGAGAAG TTTATGAATG AGGCGTTTTT
1621 ACGAATGACA GCTGCTGAGC AGGTATACAG TTTAGATACC TTTTTTTAAT TTCTTTAGCA
1681 TGATATAACT TTAGGTTTCT CATTTTAATG TATGTTGTGT GGTAGGTTGA TCTTTTTGTA
1741 GCTACCCCAA GCAATATTCC AGCAGAGTCA TTTGAAGTTT ACGAAGTTGC ACTTGCTCTT
1801 GTGGCTCAAG CTTTTATTGG TAAGAAGCCA CACCTTTTAC AGGATGCTGA TAAGCAATTC
1861 CAGCAACTTC AGCAGGCTAA GGTAATGGCT ATGGAGATTC CTGCGATGTT GTATGATACA
1921 CGGAATAATT GGGAGATAGA CTTCGGTCTA GAAAGGGGAC TCTGTGCACT GCTTATAGGC
1981 AAAGTTGATG AATGCCGTAT GTGGTTGGGC TTAGACAGTG AGGATTCACA ATATAGGAAT
2041 CCAGCTATTG TGGAGTTTGT TTTGGAGAAT TCAAATCGTG ATGACAATGA TGATCTCCCT
2101 GGACTATGCA AATTGTTGGA AACCTGGTTG GCAGGGGTTG TCTTTCCTAG GTTCAGAGAC
2161 ACCAAAGATA AAAAATTTAA ACTCGGGGAC TACTATGATG ATCCTATGGT TTTGAGTTAC
2221 TTGGAAAGAG TGGAGGTAGT TCAGGGTTCT CCTTTAGCTG CTGCTGCAAC TATGGCAAGG
2281 ATTGGAGCCG AGCATGTGAA AGCTAGTGCT ATGCAGGCAC TGCAGAAAGT TTTTCCTTCC
2341 CGCTATACAG ATAGAAACTC GGCTGAACCC AAGGATGTGC AAGAGACAGT GTTTAGTGTA
2401 GATCCTGTTG GTAACAATGT AGGCCGTGAT GGTGAGCCTG GTGTCTTTAT TGCAGAAGCT
2461 GTAAGACCCT CTGAAAACTT TGAAACTAAT GATTATGCAA TTCGAGCTGG GGTCTCAGAG
```

```
2521 AGTAGCGTTG ATGAAACTAC TGTTGAAATG TCCGTTGCTG ATATGTTAAA GGAGGCAAGT
2581 GTGAAGATCC TAGCTGCTGG TGTGGCAATT GGACTGATTT CACTGTTCAG CCAGAAGTAT
2641 TTTCTTAAAA GCAGCTCATC TTTTCAACGC AAGGATATGG TTTCTTCTAT GGAATCTGAT
2701 GTCGCTACCA TAGGTATGAT TAAATGATGC AATTTTCATA TATCTGCATT GCTCAAAATA
2761 TGCTTGTTTT GTGAGCTAAG AACATAGTTC CCACTTAATA CATGTCCCAA AAGTTGTACC
2821 AAGATTAACA AGTTGCTGAG TAAATTTCAC TAATTATGCT GCTTGAATTT TTTGATCAAA
2881 CTGTAGACAG AAATGTAAAT TTCACTCTCA ACATTTCTGT TTAGAATAAC GTAGGATTAG
2941 AGATTGCCTT AGTGTGGCTT TGTCCAACTT TTCTTTCCTT GATTTTTTTC TTTTCGATTT
3001 AGGGTCAGTC AGAGCTGACG ATTCAGAAGC ACTTCCCAGA ATGGATGCTA GGACTGCAGA
3061 GAATATAGTA TCCAAGTGGC AGAAGATTAA GTCTCTGGCT TTTGGGCCTG ATCACCGCAT
3121 AGAAATGTTA CCAGAGGTGA GGGAATAAAT CTACAATTCA ATCAATTGTG TGAAAACTGT
3181 TGGACATGAT TATAGTCTGG TGCCTTGTTT GATTCTGTTA TTTATAGGTT TTGGATGGGC
3241 GAATGCTGAA GATTTGGACT GACAGAGCAG CTGAAACTGC GCAGCTTGGG TTGGTTTATG
3301 ATTATACACT GTTGAAACTA TCTGTTGACA GTGTGACAGT CTCAGCAGAT GGAACCCGTG
3361 CTCTGGTGGA AGCAACTCTG GAGGAGTCTG CTTGTCTATC TGATTTGGTT CATCCAGAAA
3421 ACAATGCTAC TGATGTCAGA ACCTACACAA CAAGATACGA AGTTTTCTGG TCCAAGTCAG
3481 GGTGGAAAAT CACTGAAGGC TCTGTTCTTG CATCATAATA TACTCATATG TAGCATGTCT
3541 GAGCTTGCGA GATTCTCTTT GTTCTGTAAA TTCTCTCTCT AAGTTAGTGT TTATAAATGA
3601 ACACAAAAAA ATTAACGTTC TTGGCACACC CTTTTCCTTG ATCTAAACTA TAACATAAGG
3661 GCTACAA
```

C. predicted cDNA sequence of mutated AtFtn2 gene (SEQ ID NO:9) synonym: At5g42480; synonym: *ARC6*)

Sequence length = 2406 nt
Start codon (ATG) is at position 1-3
Premature stop codon (TGA) is at position 973-975
Stop codon (TAA) is at position 2404-2406

```
1    ATGGAAGCTC TGAGTCACGT CGGCATTGGT CTCTCCCCAT TCCAATTATG CCGATTACCA
61   CCGGCGACGA CAAAGCTCCG ACGTAGCCAC AACACCTCTA CAACTATCTG CTCCGCCAGC
121  AAATGGGCCG ACCGTCTTCT CTCCGACTTC AATTTCACCT CCGATTCCTC CTCCTCCTCC
181  TTCGCCACCG CCACCACCAC CGCCACTCTC GTCTCTCTGC CACCATCTAT TGATCGTCCC
241  GAACGCCACG TCCCCATCCC CATTGATTTC TACCAGGTAT TAGGAGCTCA AACACATTTC
301  TTAACCGATG GAATCAGAAG AGCATTCGAA GCTAGGGTTT CGAAACCGCC GCAATTCGGT
361  TTCAGCGACG ACGCTTTAAT CAGCCGGAGA CAGATTCTTC AAGCTGCTTG CGAAACTCTG
421  TCTAATCCTC GGTCTAGAAG AGAGTACAAT GAAGGTCTTC TTGATGATGA AGAAGCTACA
481  GTCATCACTG ATGTTCCTTG GGATAAGGTT CCTGGGGCTC TCTGTGTATT GCAAGAAGGT
541  GGTGAGACTG AGATAGTTCT TCGGGTTGGT GAGGCTCTGC TTAAGGAGAG GTTGCCTAAG
601  TCGTTTAAGC AAGATGTGGT TTTAGTTATG GCGCTTGCGT TTCTCGATGT CTCGAGGGAT
661  GCTATGGCAT TGGATCCACC TGATTTTATT ACTGGTTATG AGTTTGTTGA GGAAGCTTTG
721  AAGCTTTTAC AGGAGGAAGG AGCAAGTAGC CTTGCACCGG ATTTACGTGC ACAAATTGAT
781  GAGACTTTGG AAGAGATCAC TCCGCGTTAT GTCTTGGAGC TACTTGGCTT ACCGCTTGGT
841  GATGATTACG CTGCGAAAAG ACTAAATGGT TTAAGCGGTG TGCGGAATAT TTTGTGGTCT
901  GTTGGAGGAG GTGGAGCATC AGCTCTTGTT GGGGGTTTGA CCCGTGAGAA GTTTATGAAT
961  GAGGCGTTTT TATGAATGAC AGCTGCTGAG CAGGTTGATC TTTTTGTAGC TACCCCAAGC
1021 AATATTCCAG CAGAGTCATT TGAAGTTTAC GAAGTTGCAC TTGCTCTTGT GGCTCAAGCT
1081 TTTATTGGTA AGAAGCCACA CCTTTTACAG GATGCTGATA AGCAATTCCA GCAACTTCAG
1141 CAGGCTAAGG TAATGGCTAT GGAGATTCCT GCGATGTTGT ATGATACACG GAATAATTGG
1201 GAGATAGACT TCGGTCTAGA AAGGGGACTC TGTGCACTGC TTATAGGCAA AGTTGATGAA
1261 TGCCGTATGT GGTTGGGCTT AGACAGTGAG GATTCACAAT ATAGGAATCC AGCTATTGTG
1321 GAGTTTGTTT TGGAGAATTC AAATCGTGAT GACAATGATG ATCTCCCTGG ACTATGCAAA
1381 TTGTTGGAAA CCTGGTTGGC AGGGGTTGTC TTTCCTAGGT TCAGAGACAC CAAAGATAAA
1441 AAATTTAAAC TCGGGGACTA CTATGATGAT CCTATGGTTT TGAGTTACTT GGAAAGAGTG
1501 GAGGTAGTTC AGGGTTCTCC TTTAGCTGCT GCTGCAGCTA TGGCAAGGAT TGGAGCCGAG
1561 CATGTGAAAG CTAGTGCTAT GCAGGCACTG CAGAAAGTTT TTCCTTCCCG CTATACAGAT
1621 AGAAACTCGG CTGAACCCAA GGATGTGCAA GAGACAGTGT TTAGTGTAGA TCCTGTTGGT
1681 AACAATGTAG GCCGTGATGG TGAGCCTGGT GTCTTTATTG CAGAAGCTGT AAGACCCTCT
1741 GAAAACTTTG AAACTAATGA TTATGCAATT CGAGCTGGGG TCTCAGAGAG TAGCGTTGAT
1801 GAAACTACTG TTGAAATGTC CGTTGCTGAT ATGTTAAAGG AGGCAAGTGT GAAGATCCTA
1861 GCTGCTGGTG TGGCAATTGG ACTGATTTCA CTGTTCAGCC AGAAGTATTT TCTTAAAAGC
1921 AGCTCATCTT TTCAACGCAA GGATATGGTT TCTTCTATGG AATCTGATGT CGCTACCATA
1981 GGGTCAGTCA GAGCTGACGA TTCAGAAGCA CTTCCCAGAA TGGATGCTAG GACTGCAGAG
2041 AATATAGTAT CCAAGTGGCA GAAGATTAAG TCTCTGGCTT TTGGGCCTGA TCACCGCATA
2101 GAAATGTTAC CAGAGGTTTT GGATGGGCGA ATGCTGAAGA TTTGGACTGA CAGAGCAGCT
2161 GAAACTGCGC AGCTTGGGTT GGTTTATGAT TATACACTGT TGAAACTATC TGTTGACAGT
2221 GTGACAGTCT CAGCAGATGG AACCCGTGCT CTGGTGGAAG CAACTCTGGA GGAGTCTGCT
2281 TGTCTATCTG ATTTGGTTCA TCCAGAAAAC AATGCTACTG ATGTCAGAAC CTACACAACA
2341 AGATACGAAG TTTTCTGGTC CAAGTCAGGG TGGAAAATCA CTGAAGGCTC TGTTCTTGCA
2401 TCATAA
```

D. Genomic sequence of mutated AtFtn2 gene (SEQ ID NO:10) (synonym: At5g42480; synonym: *ARC6*)

Sequence length = 3667 nt
This sequence contains 480 nt of the 5' and 149 nt of the 3' region
Start codon (ATG) is at position 481-483
Premature stop codon (TGA) is at position 1622-1624
Stop codon (TAA) is at position 3516-3518

```
1    TGTTCTGCAT TAAGGAGAAT ACAATTATAA GCAATTTGTC TTGATTTCAA CAAGATTTTG
61   CTTGGCTATA GGATTCATTG GCTCTGTTTG CTTTTACATT TACATGTCAT AATAGTTTCG
121  AATTTTACAC ATTTCAGTTG GATGTTAAGA AAAGAGAGGG AATTGATGGG GTTTTGTGGG
181  TTTAAACTTT AAAGTAGTCA AGAATTAAGT CATTGGTTTA CTGTTGCTCT ATATGTGTAA
241  AATGAAGGCA ACTCCAACGG TTCTTAGGTG GAATAGATTA TTTAGACGAT TTAACATCAT
301  AAAGTCCGTG GCGACTGTAA CATCATAGAT TGTTTTTTAT TTTTTTCAGT AGCTGGTGAT
361  GTTTTTTGAT TTAACTTATA CTACTCAAAA TCAAAATTCC ATAAACCCTA GACGACCAAA
421  CAGTCTCTTC AATATGTAAA ACAGAACAAA GTTTTTGTAG TAGCCTAAAA AGACACTCCC
481  ATGGAAGCTC TGAGTCACGT CGGCATTGGT CTCTCCCCAT TCCAATTATG CCGATTACCA
541  CCGGCGACGA CAAAGCTCCG ACGTAGCCAC AACACCTCTA CAACTATCTG CTCCGCCAGC
601  AAATGGGCCG ACCGTCTTCT CTCCGACTTC AATTTCACCT CCGATTCCTC CTCCTCCTCC
661  TTCGCCACCG CCACCACCAC CGCCACTCTC GTCTCTCTGC CACCATCTAT TGATCGTCCC
721  GAACGCCACG TCCCCATCCC CATTGATTTC TACCAGGTAT TAGGAGCTCA AACACATTTC
781  TTAACCGATG GAATCAGAAG AGCATTCGAA GCTAGGGTTT CGAAACCGCC GCAATTCGGT
841  TTCAGCGACG ACGCTTTAAT CAGCCGGAGA CAGATTCTTC AAGCTGCTTG CGAAACTCTG
901  TCTAATCCTC GGTCTAGAAG AGAGTACAAT GAAGGTCTTC TTGATGATGA AGAAGCTACA
961  GTCATCACTG ATGTTCCTTG GGATAAGGTA ATTTCGATTT CGGAATAATA AAGTTTCTTC
1021 GTTTTAATTT CATGAATTGG ATAAAGGAAG GAACTTTTAT CTAGTGAAGG TTCCTGGGGC
1081 TCTCTGTGTA TTGCAAGAAG GTGGTGAGAC TGAGATAGTT CTTCGGGTTG GTGAGGCTCT
1141 GCTTAAGGAG AGGTTGCCTA AGTCGTTTAA GCAAGATGTG GTTTTAGTTA TGGCGCTTGC
1201 GTTTCTCGAT GTCTCGAGGG ATGCTATGGC ATTGGATCCA CCTGATTTTA TTACTGGTTA
1261 TGAGTTTGTT GAGGAAGCTT TGAAGCTTTT ACAGGTAGTT TGACTTGCTT TGGTAATTTG
1321 ACGAGCGTTG GCTTTATAAG AACTTTCTTG ATTTGATACT TTGTTATTGA GTCTTGTGTA
1381 GGAGGAAGGA GCAAGTAGCC TTGCACCGGA TTTACGTGCA CAAATTGATG AGACTTTGGA
1441 AGAGATCACT CCGCGTTATG TCTTGGAGCT ACTTGGCTTA CCGCTTGGTG ATGATTACGC
1501 TGCGAAAAGA CTAAATGGTT TAAGCGGTGT GCGGAATATT TTGTGGTCTG TTGGAGGAGG
1561 TGGAGCATCA GCTCTTGTTG GGGGTTTGAC CCGTGAGAAG TTTATGAATG AGGCGTTTTT
1621 ATGAATGACA GCTGCTGAGC AGGTATACAG TTTAGATACC TTTTTTTAAT TTCTTTAGCA
1681 TGATATAACT TTAGGTTTCT CATTTTAATG TATGTTGTGT GGTAGGTTGA TCTTTTTGTA
1741 GCTACCCCAA GCAATATTCC AGCAGAGTCA TTTGAAGTTT ACGAAGTTGC ACTTGCTCTT
1801 GTGGCTCAAG CTTTTATTGG TAAGAAGCCA CACCTTTTAC AGGATGCTGA TAAGCAATTC
1861 CAGCAACTTC AGCAGGCTAA GGTAATGGCT ATGGAGATTC CTGCGATGTT GTATGATACA
1921 CGGAATAATT GGGAGATAGA CTTCGGTCTA GAAAGGGGAC TCTGTGCACT GCTTATAGGC
1981 AAAGTTGATG AATGCCGTAT GTGGTTGGGC TTAGACAGTG AGGATTCACA ATATAGGAAT
2041 CCAGCTATTG TGGAGTTTGT TTTGGAGAAT TCAAATCGTG ATGACAATGA TGATCTCCCT
2101 GGACTATGCA AATTGTTGGA AACCTGGTTG GCAGGGGTTG TCTTTCCTAG GTTCAGAGAC
2161 ACCAAAGATA AAAAATTTAA ACTCGGGGAC TACTATGATG ATCCTATGGT TTTGAGTTAC
2221 TTGGAAAGAG TGGAGGTAGT TCAGGGTTCT CCTTTAGCTG CTGCTGCAGC TATGGCAAGG
2281 ATTGGAGCCG AGCATGTGAA AGCTAGTGCT ATGCAGGCAC TGCAGAAAGT TTTTCCTTCC
2341 CGCTATACAG ATAGAAACTC GGCTGAACCC AAGGATGTGC AAGAGACAGT GTTTAGTGTA
2401 GATCCTGTTG GTAACAATGT AGGCCGTGAT GGTGAGCCTG GTGTCTTTAT TGCAGAAGCT
2461 GTAAGACCCT CTGAAAACTT TGAAACTAAT GATTATGCAA TTCGAGCTGG GGTCTCAGAG
2521 AGTAGCGTTG ATGAAACTAC TGTTGAAATG TCCGTTGCTG ATATGTTAAA GGAGGCAAGT
```

```
2581 GTGAAGATCC TAGCTGCTGG TGTGGCAATT GGACTGATTT CACTGTTCAG CCAGAAGTAT
2641 TTTCTTAAAA GCAGCTCATC TTTTCAACGC AAGGATATGG TTTCTTCTAT GGAATCTGAT
2701 GTCGCTACCA TAGGTATGAT TAAATGATGC AATTTTCATA TATCTGCATT GCTCAAAATA
2761 TGCTTGTTTT GTGAGCTAAG AACATAGTTC CCACTTAATA CATGTCCCAA AAGTTGTACC
2821 AAGATTAACA AGTTGCTGAG TAAATTTCAC TAATTATGCT GCTTGAATTT TTTGATCAAA
2881 CTGTAGACAG AAATGTAAAT TTCACTCTCA ACATTTCTGT TTAGAATAAC GTAGGATTAG
2941 AGATTGCCTT AGTGTGGCTT TGTCCAACTT TTCTTTCCTT GATTTTTTTC TTTTCGATTT
3001 AGGGTCAGTC AGAGCTGACG ATTCAGAAGC ACTTCCCAGA ATGGATGCTA GGACTGCAGA
3061 GAATATAGTA TCCAAGTGGC AGAAGATTAA GTCTCTGGCT TTTGGGCCTG ATCACCGCAT
3121 AGAAATGTTA CCAGAGGTGA GGGAATAAAT CTACAATTCA ATCAATTGTG TGAAAACTGT
3181 TGGACATGAT TATAGTCTGG TGCCTTGTTT GATTCTGTTA TTTATAGGTT TTGGATGGGC
3241 GAATGCTGAA GATTTGGACT GACAGAGCAG CTGAAACTGC GCAGCTTGGG TTGGTTTATG
3301 ATTATACACT GTTGAAACTA TCTGTTGACA GTGTGACAGT CTCAGCAGAT GGAACCCGTG
3361 CTCTGGTGGA AGCAACTCTG GAGGAGTCTG CTTGTCTATC TGATTTGGTT CATCCAGAAA
3421 ACAATGCTAC TGATGTCAGA ACCTACACAA CAAGATACGA AGTTTTCTGG TCCAAGTCAG
3481 GGTGGAAAAT CACTGAAGGC TCTGTTCTTG CATCATAATA TACTCATATG TAGCATGTCT
3541 GAGCTTGCGA GATTCTCTTT GTTCTGTAAA TTCTCTCTCT AAGTTAGTGT TTATAAATGA
3601 ACACAAAAAA ATTAACGTTC TTGGCACACC CTTTTCCTTG ATCTAAACTA TAACATAAGG
3661 GCTACAA
```

FIG. 2

Amino Acid Sequences

A. predicted amino acid sequence of AtFtn2 (synonym: At5g42480; synonym: *ARC6*) protein

Sequence length = 801 aa

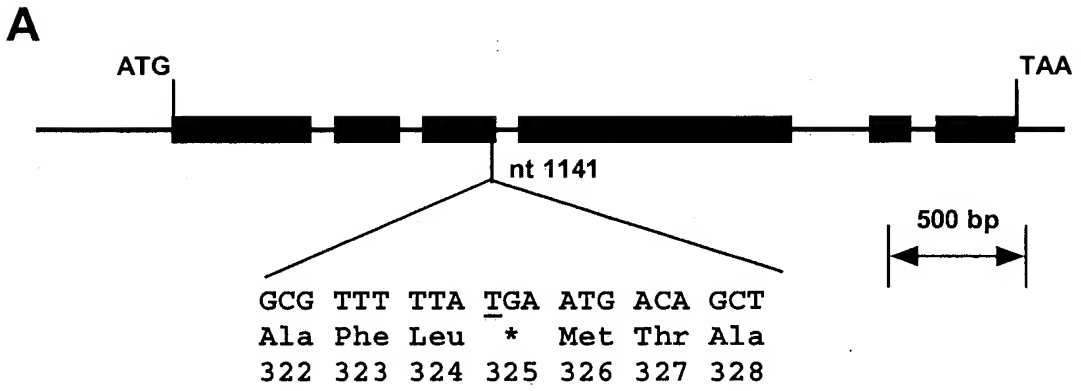
```
1   MEALSHVGIG LSPFQLCRLP PATTKLRRSH NTSTTICSAS KWADRLLSDF NFTSDSSSSS
61  FATATTTATL VSLPPSIDRP ERHVPIPIDF YQVLGAQTHF LTDGIRRAFE ARVSKPPQFG
121 FSDDALISRR QILQAACETL SNPRSRREYN EGLLDDEEAT VITDVPWDKV PGALCVLQEG
181 GETEIVLRVG EALLKERLPK SFKQDVVLVM ALAFLDVSRD AMALDPPDFI TGYEFVEEAL
241 KLLQEEGASS LAPDLRAQID ETLEEITPRY VLELLGLPLG DDYAAKRLNG LSGVRNILWS
301 VGGGGASALV GGLTREKFMN EAFLRMTAAE QVDLFVATPS NIPAESFEVY EVALALVAQA
361 FIGKKPHLLQ DADKQFQQLQ QAKVMAMEIP AMLYDTRNNW EIDFGLERGL CALLIGKVDE
421 CRMWLGLDSE DSQYRNPAIV EFVLENSNRD DNDDLPGLCK LLETWLAGVV FPRFRDTKDK
481 KFKLGDYYDD PMVLSYLERV EVVQGSPLAA AATMARIGAE HVKASAMQAL QKVFPSRYTD
541 RNSAEPKDVQ ETVFSVDPVG NNVGRDGEPG VFIAEAVRPS ENFETNDYAI RAGVSESSVD
601 ETTVEMSVAD MLKEASVKIL AAGVAIGLIS LFSQKYFLKS SSSFQRKDMV SSMESDVATI
661 GSVRADDSEA LPRMDARTAE NIVSKWQKIK SLAFGPDHRI EMLPEVLDGR MLKIWTDRAA
721 ETAQLGLVYD YTLLKLSVDS VTVSADGTRA LVEATLEESA CLSDLVHPEN NATDVRTYTT
781 RYEVFWSKSG WKITEGSVLA S*
```

B. predicted amino acid sequence of mutated AtFtn2 (synonym: At5g42480; synonym: *ARC6*) protein

Sequence length = 324 aa
The mutated protein is truncated as a result of arc6 mutation
(premature stop)

```
1   MEALSHVGIG LSPFQLCRLP PATTKLRRSH NTSTTICSAS KWADRLLSDF NFTSDSSSSS
61  FATATTTATL VSLPPSIDRP ERHVPIPIDF YQVLGAQTHF LTDGIRRAFE ARVSKPPQFG
121 FSDDALISRR QILQAACETL SNPRSRREYN EGLLDDEEAT VITDVPWDKV PGALCVLQEG
181 GETEIVLRVG EALLKERLPK SFKQDVVLVM ALAFLDVSRD AMALDPPDFI TGYEFVEEAL
241 KLLQEEGASS LAPDLRAQID ETLEEITPRY VLELLGLPLG DDYAAKRLNG LSGVRNILWS
301 VGGGGASALV GGLTREKFMN EAFL*
```

FIG. 3


A
ATG
TAA
nt 1141
500 bp
GCG TTT TTA TGA ATG ACA GCT
Ala Phe Leu * Met Thr Ala
322 323 324 325 326 327 328

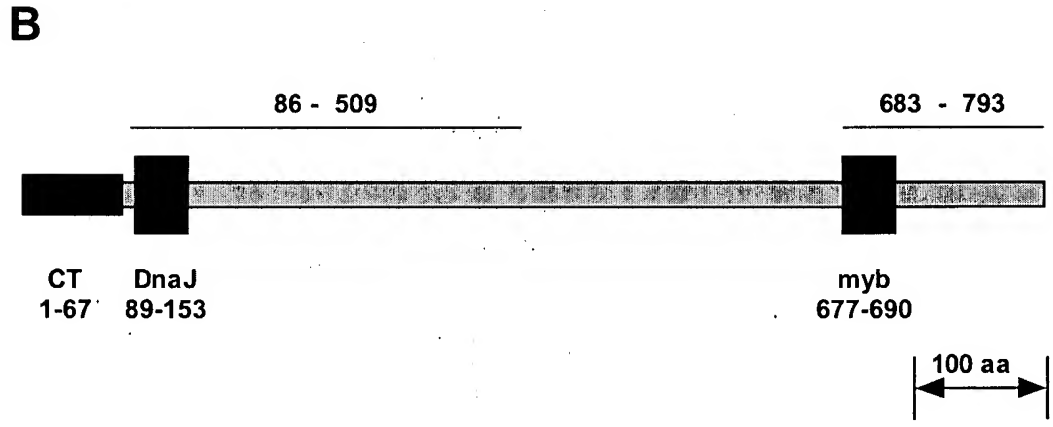

B
86 - 509
683 - 793
CT
1-67
DnaJ
89-153
myb
677-690
100 aa

FIG. 4

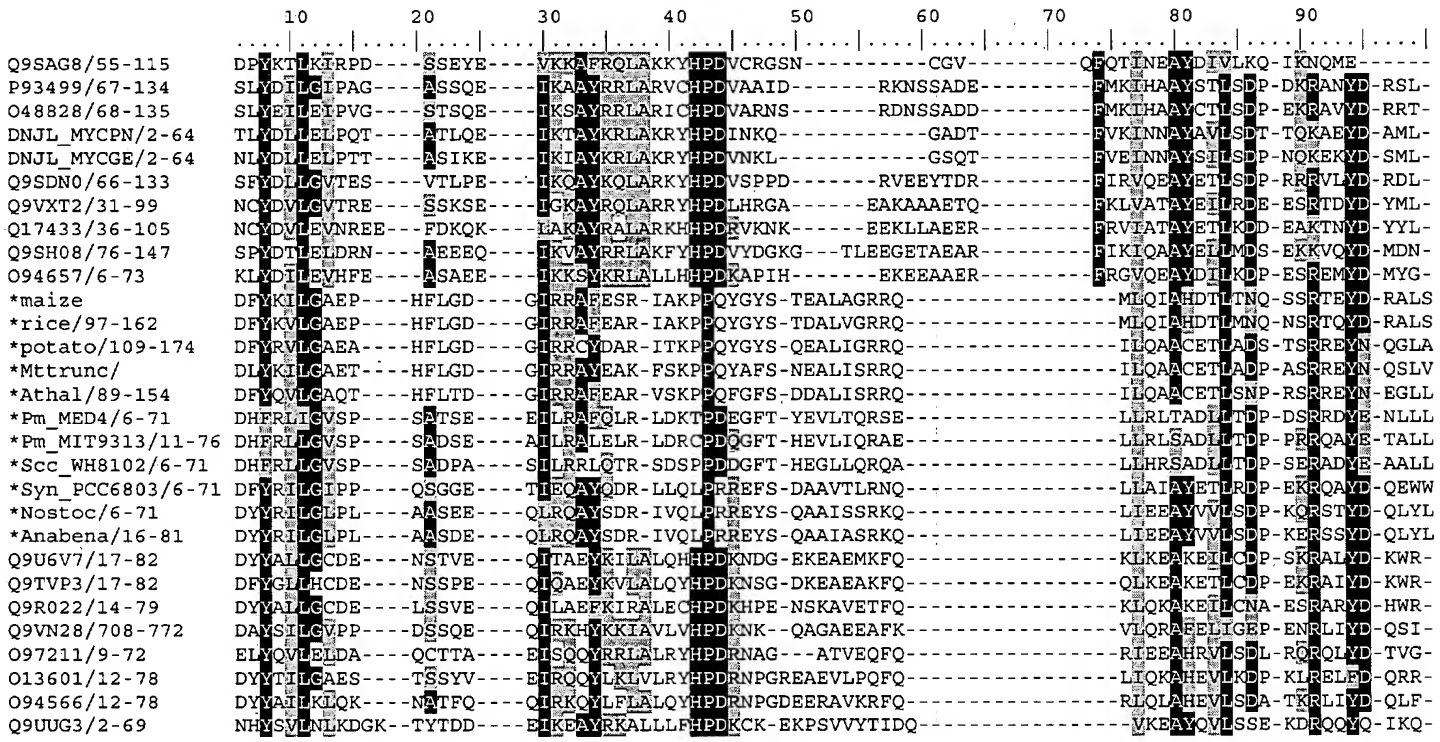

10
20
30
40
50
60
70
80
90
Q9SAG8/55-115
P93499/67-134
O48828/68-135
DNJL_MYCPN/2-64
DNJL_MYCGE/2-64
Q9SDN0/66-133
Q9VXT2/31-99
Q17433/36-105
Q9SH08/76-147
O94657/6-73
*maize
*rice/97-162
*potato/109-174
*Mttrunc/
*Athal/89-154
*Pm_MED4/6-71
*Pm_MIT9313/11-76
*Scc_WH8102/6-71
*Syn_PCC6803/6-71
*Nostoc/6-71
*Anabena/16-81
Q9U6V7/17-82
Q9TVP3/17-82
Q9R022/14-79
Q9VN28/708-772
O97211/9-72
O13601/12-78
O94566/12-78
Q9UUG3/2-69

FIG. 5

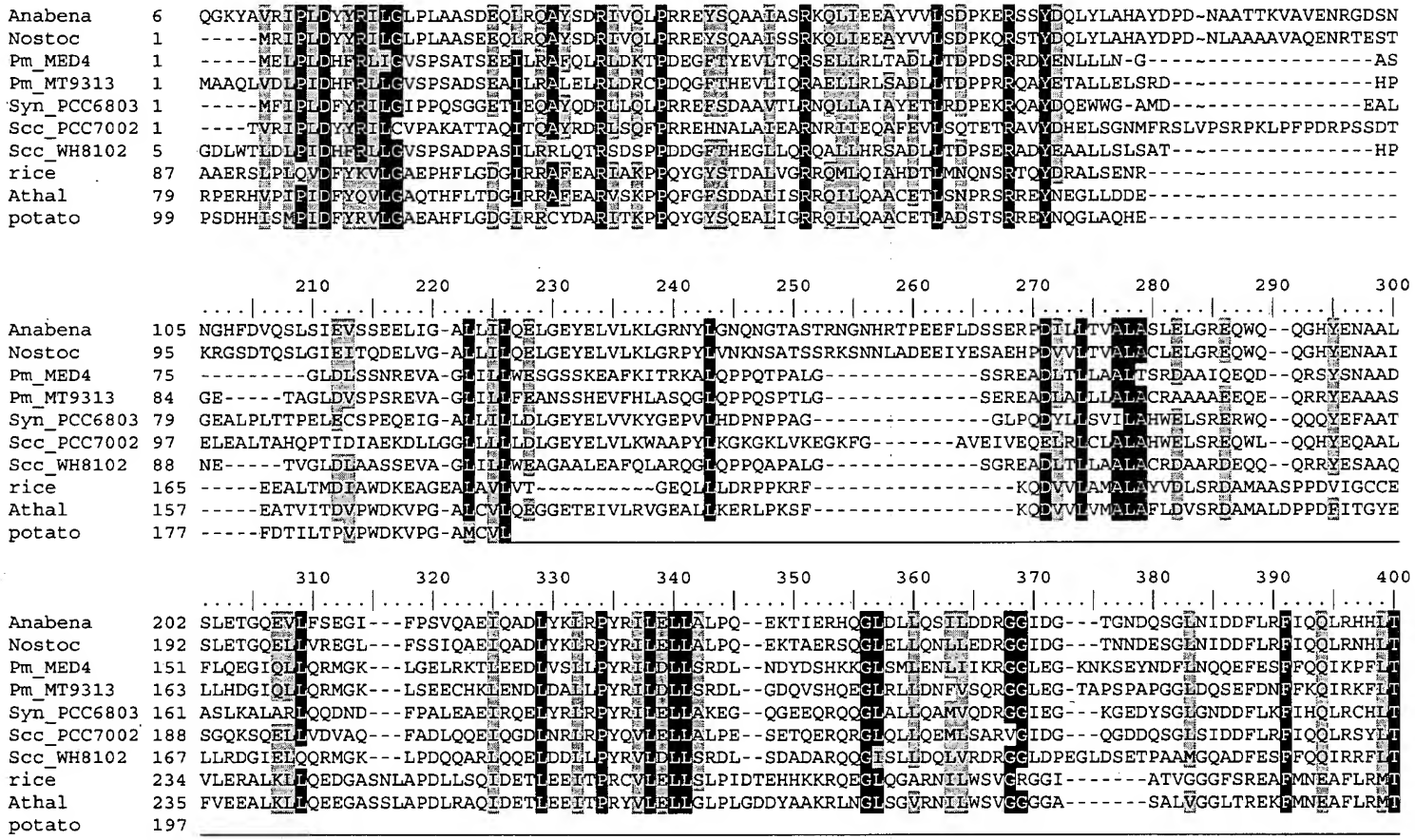

Anabena 6 QGKYAVRIPLDYYRILGLPLAASDEQLRQAYSDRIVQLPRREYSQAAIASRKQLIEEAYVVLSDPKERSSYDQLYLAHAYDPD-NAATTKVAVENRGDSN
Nostoc 1 -----MRIPLDYYRILGLPLAASEEQLRQAYSDRIVQLPRREYSQAAISSRKQLIEEAYVVLSDPKQRSTYDQLYLAHAYDPD-NLAAAAVAQENRTEST
Pm_MED4 1 -----MEIPLDHFRLIGVSPSATSEEILRAFQLRLDKTPDEGFTYEVLTQRSELLRLTADLLTDPDSRRDYENLLLN-G----------AS
Pm_MT9313 1 MAAQLVDLPIDHFRLLGVSPSADSEAILRALELRLDRCPDQGFTHEVLIQRAELLRLSADLLTDPPRRQAYETALLELSRD----------HP
Syn_PCC6803 1 -----MFIPLDFYRILGIPPQSGGETIEQAYQDRLLQLPRREFSDAAVTLRNQLLAIAYETLRDPEKRQAYDQEWWG-AMD----------EAL
Scc_PCC7002 1 ----TVRIPLDYYRILCVPAKATTAQITQAYRDRLSQFPRREHNALAIEARNRIIEQAFEVLSQTETRAVYDHELSGNMFRSLVPSRPKLPFPDRPSSDT
Scc_WH8102 5 GDLWTLDLPIDHFRLLGVSPSADPASILRRLQTRSDSPPDDGFTHEGLLQRQALLHRSADLLTDPSERADYEAALLSLSAT----------HP
rice 87 AAERSLPLQVDFYKVLGAEPHFLGDGIRRAFEARIAKPPQYGYSTDALVGRRQMLQIAHDTLMNQNSRTQYDRALSENR----------
Athal 79 RPERHVPIPIDFYQVLGAQTHFLTDGIRRAFEARVSKPPQFGFSDDALISRRQILQAACETLSNPRSRREYNEGLLDDE----------
potato 99 PSDHHISMPIDFYRVLGAEAHFLGDGIRRCYDARITKPPQYGYSQEALIGRRQILQAACETLADSTSRREYNQGLAQHE----------
210 220 230 240 250 260 270 280 290 300
Anabena 105 NGHFDVQSLSIEVSSEELIG-ALLILQELGEYELVLKLGRNYLGNQNGTASTRNGNHRTPEEFLDSSERPDILLTVALASLELGREQWQ--QGHYENAAL
Nostoc 95 KRGSDTQSLGIEITQDELVG-ALLILQELGEYELVLKLGRPYLVNKNSATSSRKSNNLADEEIYESAEHPDVVLTVALACLELGREQWQ--QGHYENAAI
Pm_MED4 75 ---------GLDLSSNREVA-GLILLWESGSSKEAFKITRKALQPPQTPALG----------SSREADLTLLAALTSRDAAIQEQD--QRSYSNAAD
Pm_MT9313 84 GE-----TAGLDVSPSREVA-GLILLFEANSSHEVFHLASQGLQPPQSPTLG----------SEREADLALLLALACRAAAAEEQE--QRRYEAAAS
Syn_PCC6803 79 GEALPLTTPELECSPEQEIG-ALLILLDLGEYELVVKYGEPVLHDPNPPAG----------GLPQDYLLSVILAHWELSRERWQ--QQQYEFAAT
Scc_PCC7002 97 ELEALTAHQPTIDIAEKDLLGGLLILLDLGEYELVLKWAAPYLKGKGKLVKEGKFG-------AVEIVEQELRLCLALAHWELSREQWL--QQHYEQAAL
Scc_WH8102 88 NE-----TVGLDLAASSEVA-GLILLWEAGAALEAFQLARQGLQPPQAPALG----------SGREADLTLLAALACRDAARDEQQ--QRRYESAAQ
rice 165 -----EEALTMDIAWDKEAGEALAVLVT----------GEQLLLDRPPKRF----------KQDVVLAMALAYVDLSRDAMAASPPDVIGCCE
Athal 157 -----EATVITDVPWDKVPG-ALCVLQEGGETEIVLRVGEALLKERLPKSF----------KQDVVLVMALAFLDVSRDAMALDPPDEITGYE
potato 177 -----FDTILTPVPWDKVPG-AMCVL
310 320 330 340 350 360 370 380 390 400
Anabena 202 SLETGQEVLFSEGI---FPSVQAEIQADLYKLRPYRILELLALPQ--EKTIERHQGLDLLQSILDDRGGIDG---TGNDQSGLNIDDFLRFIQQLRHHLT
Nostoc 192 SLETGQELLVREGL---FSSIQAEIQADLYKLRPYRILELLALPQ--EKTAERSQGLELLQNLLEDRGGIDG---TNNDESGLNIDDFLRFIQQLRNHLT
Pm_MED4 151 FLQEGIQLLQRMGK---LGELRKTLEEDLVSLLPYRILDLLSRDL--NDYDSHKKGLSMLENLIIKRGGLEG-KNKSEYNDFLNQQEFESFFQQIKPFLT
Pm_MT9313 163 LLHDGIQLLQRMGK---LSEECHKLENDLDALLPYRILDLLSRDL--GDQVSHQEGLRLLDNFVSQRGGLEG-TAPSPAPGGLDQSEFDNFFKQIRKFLT
Syn_PCC6803 161 ASLKALARLQQDND---FPALEAETRQELYRLRPYRILELLAKEG--QGEEQRQQGLALLQAMVQDRGGIEG---KGEDYSGLGNDDFLKFIHQLRCHLT
Scc_PCC7002 188 SGQKSQELLVDVAQ---FADLQQELQGDLNRLRPYQVLELLALPE--SETQERQRGLQLLQEMLSARVGIDG---QGDDQSGLSIDDFLRFIQQLRSYLT
Scc_WH8102 167 LLRDGIELQQRMGK---LPDQQARLQQELDDLLPYRVLDLLSRDL--SDADARQQGISLLDQLVRDRGGLDPEGLDSETPAAMGQADFESFFQQIRRFLT
rice 234 VLERALKLLQEDGASNLAPDLLSQIDETLEEITPRCVLELLSLPIDTEHHKKRQEGLQGARNILWSVGRGGI-------ATVGGGFSREAFMNEAFLRMT
Athal 235 FVEEALKLLQEEGASSLAPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNILWSVGGGGA-------SALVGGLTREKFMNEAFLRMT
potato 197

FIG. 5, continued (2/3)

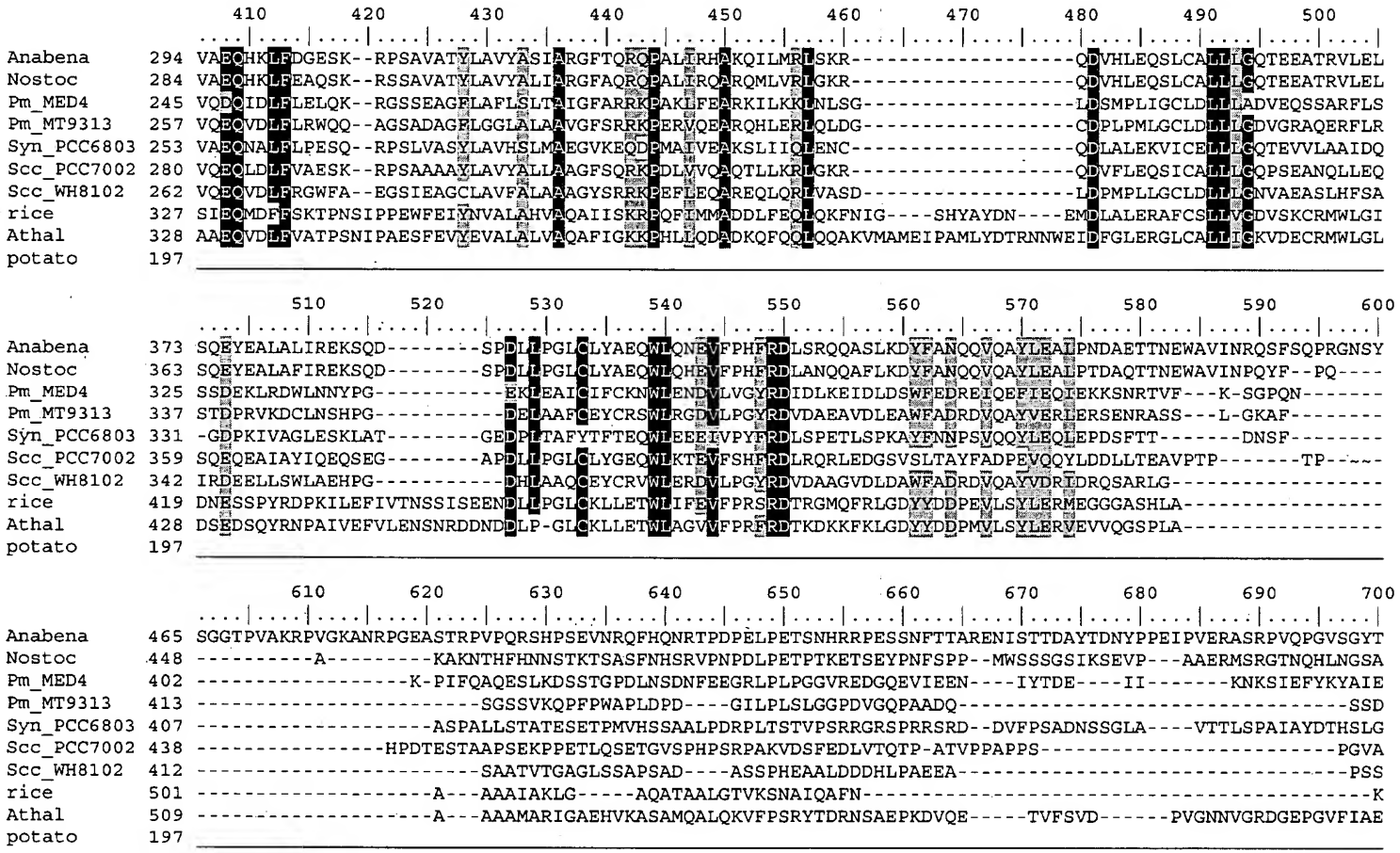


FIG. 5, continued (3/3)

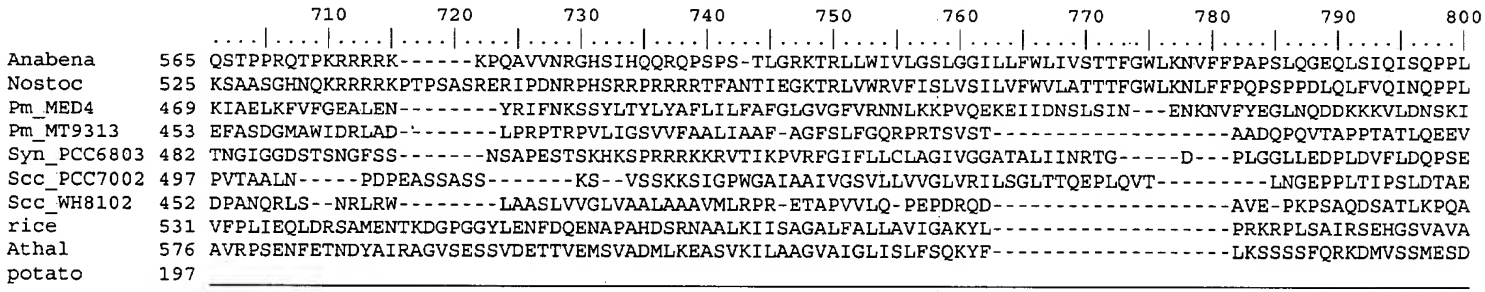


FIG. 6

Synechococcus sp. PCC 7942 cell division protein Ftn2 gene

A. Ftn2 DNA nucleic acid sequence (SEQ ID NO:4)

1 cttgccgact aaaggctaag catcgccatt ccttagatta aagcagtctg tcggcggcgc
61 tgtgccggtt aacaccagtc tgtcgctgac agcggtgcct ttctggggct tgcctgtggg
121 gcgagtaacc gatcgctggg ataagagttg gtgcttctgg ctctcaagaa tagggttttc
181 cgtcgcgtat tcccgatcac atccccctgt gtctgctacg gagataacgc cgatcactca
241 acagaattgg taagttgacg gtcaagttgg gatgatgaag tcggctcaag ctggcgatcc
301 ggatctggtg ggtgttctgt gcgtattcct ctcgattact accgaattct ctgtgttggc
361 gtgcaagcct cggcagacaa acttgccgaa agctaccgcg atcgcctcaa ccaatcgccc
421 tcccatgagt tttcagagct ggcattgcag gcgcggcggc aactcctcga agcagcgatt
481 gctgagctga gtgatcccga acagcgcgat cgctacgatc gccgcttttt tcagggcggt
541 ctggaagcga ttgaaccaag cctagaactc gaagactggc agcgaattgg agccctgctg
601 atcctgctgg aattggggga atacgatcgc gtttcgcaac tggctgagga actcctgcca
661 gactacgacg cgagcgcaga agtacgcgat cagttcgcgc ggggtgatat cgccttggcg
721 atcgcactat cccagcaatc cctcggtcga gaatgccgtc agcagggtct gtacgaacag
781 gccgcccagc actttggccg cagccagtct gccctagccg atcatcagcg ctttcctgaa
841 ctgagtcgaa ccctgcacca agaacaagga cagctacggc cctatcgcat tttggagcgg
901 ttggcccagc ccttgactgc cgatagcgat cgccagcagg gtttgctgtt gttgcaggcg
961 atgttggacg accggcaggg cattgaaggc cctggggatg atggctcggg gctgaccctt
1021 gataactttt tgatgtttct ccagcaaatt cgcggctatc tgaccctggc tgaacagcag
1081 ttgctgtttg aatcggaagc gcgtcggccc tcgccggctg cgagcttttt tgcctgctac
1141 accctgattg cgcggggctt ttgcgatcac caaccctcgt tgatccatcg cgccagcttg
1201 ctcttgcatg aactcaagag ccgcatggat gtgcacatcg aacaggcgat cgccagccta
1261 ttgctcggac agcccgaaga agctgaggcg ctactcgtcc agagccaaga tgaggaaacc
1321 ctcagccaaa tccgtgccct agcccaaggg gaagccctga tcgtcggttt gtgccgattc
1381 acggaaacct ggctagcgac caaggtattt ccggatttcc gcgacctcaa ggaaaggact
1441 gcgccgctgc agccctactt tgacgacccc gatgtccaga cctatctgga tgcgatcgtg
1501 gagttgccgt ccgatttgat gccaacgccg ctacccgttg agccgcttga ggtgcgatcg
1561 tcgttgctgg ccaaggaact gccgacccca gcaacgcctg gtgtagctcc accccctcgc
1621 cgccgtcgcc gcgatcgctc cgaacgtcct gctcgcacgg ccaaacgctt gcccttgccc
1681 tggattggtt tgggggttgt ggtggttctc ggcggtggaa caggggtttg ggcttggcga
1741 tcgcgttcca attccacccc gccgaccccg ccccccgtgg ttcaaacgct gcctgaggcg

FIG. 6 continued (2/2)

1801 gtacctgccc cttcgcccgc gccagttacc gttgccctcg atcgggctca ggctgaaact
1861 gtgttgcaaa actggttggc cgctaaagct gcagccttgg ggcctcaata cgatcgcgat
1921 cgcttagcga cggtgctgac cggtgaggtt ctgcagactt ggcagggttt ttctagccag
1981 caggccaaca cccagctcac atcacagttc gatcacaagt taaccgtcga ctcagttcag
2041 ctcagtgacg gtgatcaacg agcagtagtc caagccaagg tcgatgaagt tgagcaggtc
2101 tatcgaggcg accagctgct cgaaacgcgc cgagatttgg gcttggtgat ccgctaccag
2161 ctcgtgcgcg agaacaacat ctggaaaatt gcttcgatta gtttggtgcg ctaggaattc
2221 gcaaggggtg aacccccctgc ggtcttttct gtagatcccc tagagcgatc gcagaatgtt
2281 cagcgattcc tggatgtgcg cttgggcatt caagagtgaa tcaaaaatgt ggcgcacctt
2341 gccctctttg tcgatcacat aagtgacgcg acccggaatc acaaacaggg ttttgggcac
2401 gccataggtt tgacggaggc gatcgcctgc atcgctcagc agttggaagg gcaagttgta
2461 tttctgggc

B. Ftn2 Protein amino acid sequence (SEQ ID NO:5)

translation="MRIPLDYYRILCVGVQASADKLAESYRDRLNQSPSHEFSELALQ

ARRQLLEAAIAELSDPEQRDRYDRRFFQGGLEAIEPSLELEDWQRIGALLILLELGEY

DRVSQLAEELLPDYDASAEVRDQFARGDIALAIALSQQSLGRECRQQGLYEQAAQHFG

RSQSALADHQRFPELSRTLHQEQGQLRPYRILERLAQPLTADSDRQQGLLLLQAMLDD

RQGIEGPGDDGSGLTLDNFLMFLQQIRGYLTLAEQQLLFESEARRPSPAASFFACYTL

IARGFCDHQPSLIHRASLLLHELKSRMDVHIEQAIASLLLGQPEEAEALLVQSQDEET

LSQIRALAQGEALIVGLCRFTETWLATKVFPDFRDLKERTAPLQPYFDDPDVQTYLDA

IVELPSDLMPTPLPVEPLEVRSSLLAKELPTPATPGVAPPPRRRRRDRSERPARTAKR

LPLPWIGLGVVVVLGGGTGVWAWRSRSNSTPPTPPPVVQTLPEAVPAPSPAPVTVALD

RAQAETVLQNWLAAKAAALGPQYDRDRLATVLTGEVLQTWQGFSSQQANTQLTSQFD

HKLTVDSVQLSDGDQRAVVQAKVDEVEQVYRGDQLLETRRDLGLVIRYQLVRENNIW

KIASISLVR"

FIG. 7

Synechococcus sp. PCC 7942 cell division protein Ftn6 gene

A. Ftn6 DNA nucleic acid sequence (SEQ ID NO:6)

```
   1 ctcgatactt gggagttgaa cacagagtag tagtctaagt aacaactgct cgtgagcaat
  61 ttgctacact ttttaccaaa ttttgagctc agttttcgcg aaaactggga tgttgagttg
 121 aaccctcagc agcaaaattg taccgcctga gacttttacc gttttattcg gccatctggg
 181 aacaatcgcc ctggagctta ttgtgacctc tacccgtact gccgttattg ccttgttaga
 241 acgctatttc gagctgtcgg cagcgcgagc agcagaggtc ttgcagcaac tgcgatcgca
 301 ccaccctgaa gcctggattt atcccgccac agtcgaggcg atttaccaag gccgttaccg
 361 ctgggtgtcg atcgcacaaa tccttgctct gtggcagcgg cgcgggcaga tcaactgcca
 421 cttcagtgca gactatgagc gcttgttgct cggtgaagtt ccagagcaac ccgatcgcat
 481 caatgttgag acgcggctcc ctgcgatcgc catgaccttg ccttgggtgc cagaacagcc
 541 tggagaagca ttcgtgccag cgcaagatca gtcgggttta actgagcgcc tttataaaac
 601 gttggtcaaa gcgggcagcg attgcgctgg gtaggcttag aacagttgcc atccaaactt
 661 gagagtgccc gttcggccag ccaagagaat tccaagagcc tttcagaacg gacaacaatt
 721 ctgctctaca atcaagcccg agtgaagagg cggcgggcta ttggctgaat ggcaaaaaac
 781 atcattcttt cagcaatcgt gggttatacc tacgacaaaa ttgacctatt cttaacttct
 841 gcactccgta acacctcagc agatattctt ttaattgcat caagtccttc agcccaactc
 901 cgtcatcagt tattgagttc acctcgggtc aaactcgttg atgtgaacct tcaaggtgaa
 961 ccagctgaaa tggtatttcg ccgtttcttt attgccaagg agattttggc gagaatcgaa
1021 gcagatgaaa ttctcttgag cgatgctcgc gatgtctatt tccaatctga ccctttggt
1081 gtccaagggg ttttatttgc cgaggaacct cagctaatcg caaactgtaa agtcaatagc
1141 agctggataa aaaaatactt aggagaggat gagtttcaag ccatttctcc taatccaatt
1201 ctctgcgggg gcaaccatgt gctggatgcc accaaggcct ttagcctgac gttgaccaca
1261 ccagaagaaa ttgttgggct gcccgagagt ttgctggcct tggcggctca agctgctcaa
1321 gccgctggtg aaacagaggc aacacccgaa gccggccctt ggcgaatcac cctcgacttc
1381 ccaagctttg
```

B. Ftn6 Protein amino acid sequence (SEQ ID NO:7)

MGTIALELIVTSTRTAVIALLERYFELSAARAAEVLQQLRSHHP

EAWIYPATVEAIYQGRYRWVSIAQILALWQRRGQINCHFSADYERLLLGEVPEQP

DRINVETRLPAIAMTLPWVPEQPGEAFVPAQDQSGLTERLYKTLVKAGSDCAG

FIG. 8

Additional Sequences

First Set

```
LOCUS       BK000999          2283 bp    mRNA    linear   PLN 06-JAN-2003
DEFINITION  CDS for rice Arc6 orthologue, predicted from AAAA01000502.
ACCESSION   BK000999
VERSION
KEYWORDS    .
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Oryza.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Vitha,S., Koksharova,O., van Erp,H., Froehlich,J.E. and
            Osteryoung,K.W.
  TITLE     Arabidopsis Arc6: A J-Domain Plastid Division Protein Whose
            Prokaryotic Ancestors Are Unique to Cyanobacteria
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Vitha,S., Koksharova,O., van Erp,H., Froehlich,J.E. and
            Osteryoung,K.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2003) Department of Plant Biology, Michigan
State
            University, 166 Plant Biology Building, East Lansing, MI 48824,
USA
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Oryza sativa"
                     /strain="indica cultivar-group"
     gene            1..2283
                     /gene="Arc6"
                     /note="Orthologue of Arabidopsis At5g42480  (Arc6)"
     CDS             1..2283
                     /gene="Arc6"
                     /note="has chloroplast targeting N-terminal signal,
                     followed by J domain"
                     /codon_start=1
                     /product="Arc6"

/translation="MEGFHNLLARPNSAPFAFSLPRPRPRPRRRPPPHPSAACRAASR

WAERLFADFHLLPTAAPSDPPSPAPAPAAAPSASPFVPLFPDAAERSLPLQVDFYKVL

GAEPHFLGDGIRRAFEARIAKPPQYGYSTDALVGRRQMLQIAHDTLMNQNSRTQYDRA

LSENREEALTMDIAWDKEAGEALAVLVTGEQLLLDRPPKRFKQDVVLAMALAYVDLSR
```

DAMAASPPDVIGCCEVLERALKLLQEDGASNLAPDLLSQIDETLEEITPRCVLELLSL

PIDTEHHKKRQEGLQGARNILWSVGRGGIATVGGGFSREAFMNEAFLRMTSIEQMDFF

SKTPNSIPPEWFEIYNVALAHVAQAIISKRPQFIMMADDLFEQLQKFNIGSHYAYDNE

MDLALERAFCSLLVGDVSKCRMWLGIDNESSPYRDPKILEFIVTNSSISEENDLLPGL

CKLLETWLIFEVFPRSRDTRGMQFRLGDYYDDPEVLSYLERMEGGGASHLAAAAAIAK

LGAQATAALGTVKSNAIQAFNKVFPLIEQLDRSAMENTKDGPGGYLENFDQENAPAHD

SRNAALKIISAGALFALLAVIGAKYLPRKRPLSAIRSEHGSVAVANSVDSTDDPALDE

DPVHIPRMDAKLAEDIVRKWQSIKSKALGPEHSVASLQEVLDGNMLKVWTDRAAEIER

HGWFWEYTLSDVTIDSITISLDGRRATVEATIDEAGQLTDVTEPRNNDSYDTKYTTRY
EMAFSKLGGWKITEGAVLKS"

BASE COUNT 551 a 576 c 592 g 564 t
ORIGIN

```
       1 atggagggct tccacaacct cctcgcccgc cccaactcgg cgccattcgc cttctccctc
      61 cctcgcccgc gcccgcgccc gcgccgcagg ccgccgcctc acccctccgc tgcctgccgc
     121 gccgcgagcc gctgggccga acgcctcttc gccgacttcc acctcctccc caccgccgcg
     181 ccctccgacc cgccgtcccc ggccccggcc ccggccgccg cgccctccgc ctccccttc
     241 gtcccgctct tccccgacgc cgccgaacgc tccctcccgc tccaagtcga tttctacaag
     301 gttctagggg cagagccaca tttccttggc gatggcatca ggagggcgtt cgaggcacgg
     361 atagccaagc caccgcagta tggctacagc acggatgctc ttgttggtcg tcgacaaatg
     421 ctgcagattg cccatgacac tctcatgaac cagaactccc gcactcagta tgatcgtgcg
     481 ctttctgaga accgtgaaga agctctcacc atggatattg cttgggacaa ggaggctggg
     541 gaggcacttg ctgtgcttgt aactggagaa cagttgcttc tggatcggcc acccaagcgc
     601 ttcaagcagg acgtggtgct agcgatggct ctggcttatg tggatctatc aagggatgct
     661 atggcagcaa gccctccaga tgtaattggc tgctgcgagg tgctcgagag ggctctcaag
     721 ctcttgcagg aagatggagc aagcaatctc gcacctgatc tgctttcaca gattgatgaa
     781 actctcgagg agattacacc tcgctgtgta ttggagcttc tctcccttcc tattgacaca
     841 gagcatcata agaagcgcca agaagggctt caaggtgcga gaaacatttt gtggagcgtt
     901 ggcagaggag gtattgctac cgttggagga ggattttctc gtgaagcctt catgaacgag
     961 gctttttga ggatgacatc aattgaacag atggatttct tttcaaaaac accgaatagc
    1021 attcctcctg aatggtttga aatttacaat gtagcacttg cacatgtcgc tcaagcaatt
    1081 ataagtaaaa ggccacaatt catcatgatg gcggatgatc tttttgaaca actccagaag
    1141 ttcaacatag gttctcatta tgcttatgat aatgagatgg accttgcatt ggaaagggca
    1201 ttctgctcat tgctagtcgg agatgttagc aagtgcagaa tgtggcttgg aattgataat
    1261 gagtcttcac catacagaga ccccaaaatt ctagagttta ttgtgaccaa ctctagcatc
    1321 agtgaagaga atgatcttct tccagggctg tgcaagcttt tggagacttg gcttatcttt
    1381 gaggtttttc ctaggagcag agatactcgg ggcatgcagt tcagacttgg agattactac
    1441 gatgatccag aagttttaag ctacctagaa aggatggagg gtggtggtgc ttctcatttg
    1501 gctgctgctg ctgctattgc aaaacttggt gctcaagcta cagctgcact tggtactgtg
    1561 aaatcaaatg ctattcaagc gttcaacaag gtttttccat tgatagaaca gttagacagg
    1621 tcagccatgg aaaatactaa agatggccct gggggatatc ttgaaaattt tgaccaggaa
    1681 aatgcacctg ctcatgattc gagaaatgcc gccttgaaga ttatctctgc tggcgcactg
    1741 tttgcactgt tggcagtaat tggggccaaa tatttgcctc gtaagaggcc cctttctgct
    1801 attaggagtg agcatggatc tgtggcagtt gctaatagtg tcgactctac tgatgatcct
    1861 gcactagatg aagatccagt acatattcct agaatggatg cgaagctggc agaagatatt
    1921 gttcgcaagt ggcagagtat caaatctaag gccttgggac cagaacattc ggttgcatca
```

```
1981 ttgcaagagg ttcttgatgg caacatgcta aaggtgtgga ctgaccgagc agcggagatt
2041 gagcgtcatg ggtggttctg ggagtataca ctatccgatg tgacgattga tagcatcact
2101 atctccctag atggtcgacg agcgactgtg gaggctacga ttgatgaggc aggccaactt
2161 actgatgtta ctgagcccag aaacaatgat tcatatgaca caaaatacac tacccggtat
2221 gagatggcct tctccaagct aggagggtgg aagataacgg aaggagcagt cctcaagtcg
2281 tag
```

```
LOCUS       BAB10489                801 aa            linear   PLN 27-DEC-
2000
DEFINITION  gene_id:MDH9.18~pir||S76082~similar to unknown protein
[Arabidopsis
            thaliana].
ACCESSION   BAB10489
VERSION     BAB10489.1  GI:9759484
DBSOURCE    locus AB016888 accession AB016888.1
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Asamizu,E., Sato,S., Kaneko,T., Nakamura,Y., Kotani,H.,
Miyajima,N.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
            Sequence features of the regions of 1,081,958 bp covered by
            seventeen physically assigned P1 and TAC clones
  JOURNAL   DNA Res. 5 (6), 379-391 (1998)
  MEDLINE   99156233
   PUBMED   10048488
REFERENCE   2  (residues 1 to 801)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-1998) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please
see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MDH9
            Genes with similarity to proteins in the databases are
described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-
081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St.
Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
```

```
            This sequence may not be the entire insert of this clone. It
may be
            shorter because we remove overlaps between neighboring
submissions.
            The 5' clone is K5J14 and the 3' clone is K16E1.
FEATURES             Location/Qualifiers
     source          1..801
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MDH9"
                     /clone_lib="Mitsui P1"
     Protein         1..801
                     /name="gene_id:MDH9.18
                     pir||S76082
                     similar to unknown protein"
     CDS             1..801
                     /coded_by="join(AB016888.1:64077..64583,
                     AB016888.1:64666..64890,AB016888.1:64978..65238,
                     AB016888.1:65322..66309,AB016888.1:66599..66732,
                     AB016888.1:66824..67114)"
ORIGIN
        1 mealshvgig lspfqlcrlp pattklrrsh ntstticsas kwadrllsdf nftsdsssss
       61 fatatttatl vspppsidrp erhvpipidf yqvlgaqthf ltdgirrafe arvskppqfg
      121 fsddalisrr qilqaacetl snprsrreyn eglldeeat vitdvpwdkv pgalcvlqeg
      181 geteivlrvg eallkerlpk sfkqdvvlvm alafldvsrd amaldppdfi tgyefveeal
      241 kllqeegass lapdlraqid etleeitpry vlellglplg ddyaakrlng lsgvrnilws
      301 vggggasalv ggltrekfmn eaflrmtaae qvdlfvatps nipaesfevy evalalvaqa
      361 figkkphllq dadkqfqqlq qakvmameip amlydtrnnw eidfglergl calligkvde
      421 crmwlgldse dsqyrnpaiv efvlensnrd dnddlpglck lletwlagvv fprfrdtkdk
      481 kfklgdyydd pmvlsylerv evvqgsplaa aaamarigae hvkasamqal qkvfpsrytd
      541 rnsaepkdvq etvfsvdpvg nnvgrdgepg vfiaeavrps enfetndyai ragvsessvd
      601 ettvemsvad mlkeasvkil aagvaiglis lfsqkyflks sssfqrkdmv ssmesdvati
      661 gsvraddsea lprmdartae nivskwqkik slafgpdhri emlpevldgr mlkiwtdraa
      721 etaqlglvyd ytllklsvds vtvsadgtra lveatleesa clsdlvhpen natdvrtytt
      781 ryevfwsksg wkitegsvla s
//
```

>gi|18422214|ref|NM_123613.1| Arabidopsis thaliana putative protein, predicted mRNA

ATGGAAGCTCTGAGTCACGTCGGCATTGGTCTCTCCCCATTCCAATTATGCCGATTACCACCGGCGACGA
CAAAGCTCCGACGTAGCCACAACACCTCTACAACTATCTGCTCCGCCAGCAAATGGGCCGACCGTCTTCT
CTCCGACTTCAATTTCACCTCCGATTCCTCCTCCTCCTCCTTCGCCACCGCCACCACCACCGCCACTCTC
GTCTCTCCGCCACCATCTATTGATCGTCCCGAACGCCACGTCCCCATCCCCATTGATTTCTACCAGGTAT
TAGGAGCTCAAACACATTTCTTAACCGATGGAATCAGAAGAGCATTCGAAGCTAGGGTTTCGAAACCGCC
GCAATTCGGTTTCAGCGACGACGCTTTAATCAGCCGGAGACAGATTCTTCAAGCTGCTTGCGAAACTCTG
TCTAATCCTCGGTCTAGAAGAGAGTACAATGAAGGTCTTCTTGATGATGAAGAAGCTACAGTCATCACTG
ATGTTCCTTGGGATAAGGTTCCTGGTGCTCTCTGTGTATTGCAAGAAGGTGGTGAGACTGAGATAGTTCT
TCGGGTTGGTGAGGCTCTGCTTAAGGAGAGGTTGCCTAAGTCGTTTAAGCAAGATGTGGTTTTAGTTATG
GCGCTTGCGTTTCTCGATGTCTCGAGGGATGCTATGGCATTGGATCCACCTGATTTTATAACTGGTTATG
AGTTTGTTGAGGAAGCTTTGAAGCTTTTACAGGAGGAAGGAGCAAGTAGCCTTGCACCGGATTTACGTGC
ACAAATTGATGAGACTTTGGAAGAGATCACTCCGCGTTATGTCTTGGAGCTACTTGGCTTACCGCTTGGT
GATGATTACGCTGCGAAAAGACTAAATGGTTTAAGCGGTGTGCGGAATATTTTGTGGTCTGTTGGAGGAG
GTGGAGCATCAGCTCTTGTTGGGGGTTTGACCCGTGAGAAGTTTATGAATGAGGCGTTTTTACGAATGAC
AGCTGCTGAGCAGGTTGATCTTTTTGTAGCTACCCCAAGCAATATTCCAGCAGAGTCATTTGAAGTTTAC
GAAGTTGCACTTGCTCTTGTGGCTCAAGCTTTTATTGGTAAGAAGCCACACCTTTTACAGGATGCTGATA
AGCAATTCCAGCAACTTCAGCAGGCTAAGGTAATGGCTATGGAGATTCCTGCGATGTTGTATGATACACG
GAATAATTGGGAGATAGACTTCGGTCTAGAAAGGGGACTCTGTGCACTGCTTATAGGCAAAGTTGATGAA
TGCCGTATGTGGTTGGGCTTAGACAGTGAGGATTCACAATATAGGAATCCAGCTATTGTGGAGTTTGTTT
TGGAGAATTCAAATCGTGATGACAATGATGATCTCCCTGGACTATGCAAATTGTTGGAAACCTGGTTGGC
AGGGGTTGTCTTTCCTAGGTTCAGAGACACCAAAGATAAAAAATTTAAACTCGGGGACTACTATGATGAT
CCTATGGTTTTGAGTTACTTGGAAAGAGTGGAGGTAGTTCAGGGTTCTCCTTTAGCTGCTGCTGCAGCTA
TGGCAAGGATTGGAGCCGAGCATGTGAAAGCTAGTGCTATGCAGGCACTGCAGAAAGTTTTTCCTTCCCG
CTATACAGATAGAAACTCGGCTGAACCCAAGGATGTGCAAGAGACAGTGTTTAGTGTAGATCCTGTTGGT
AACAATGTAGGCCGTGATGGTGAGCCTGGTGTCTTTATTGCAGAAGCTGTAAGACCCTCTGAAAACTTTG
AAACTAATGATTATGCAATTCGAGCTGGGGTCTCAGAGAGTAGCGTTGATGAAACTACTGTTGAAATGTC
CGTTGCTGATATGTTAAAGGAGGCAAGTGTGAAGATCCTAGCTGCTGGTGTGGCAATTGGACTGATTTCA
CTGTTCAGCCAGAAGTATTTTCTTAAAAGCAGCTCATCTTTTCAACGCAAGGATATGGTTTCTTCTATGG
AATCTGATGTCGCTACCATAGGGTCAGTCAGAGCTGACGATTCAGAAGCACTTCCCAGAATGGATGCTAG
GACTGCAGAGAATATAGTATCCAAGTGGCAGAAGATTAAGTCTCTGGCTTTTGGGCCTGATCACCGCATA
GAAATGTTACCAGAGGTTTTGGATGGGCGAATGCTGAAGATTTGGACTGACAGAGCAGCTGAAACTGCGC
AGCTTGGGTTGGTTTATGATTATACACTGTTGAAACTATCTGTTGACAGTGTGACAGTCTCAGCAGATGG
AACCCGTGCTCTGGTGGAAGCAACTCTGGAGGAGTCTGCTTGTCTATCTGATTTGGTTCATCCAGAAAAC
AATGCTACTGATGTCAGAACCTACACAACAAGATACGAAGTTTTCTGGTCCAAGTCAGGGTGGAAAATCA
CTGAAGGCTCTGTTCTTGCATCATAA

>gi|15238978|ref|NP_199063.1| putative protein [Arabidopsis thaliana]

MEALSHVGIGLSPFQLCRLPPATTKLRRSHNTSTTICSASKWADRLLSDFNFTSDSSSSSFATATTTATL
VSPPPSIDRPERHVPIPIDFYQVLGAQTHFLTDGIRRAFEARVSKPPQFGFSDDALISRRQILQAACETL
SNPRSRREYNEGLLDDEEATVITDVPWDKVPGALCVLQEGGETEIVLRVGEALLKERLPKSFKQDVVLVM
ALAFLDVSRDAMALDPPDFITGYEFVEEALKLLQEEGASSLAPDLRAQIDETLEEITPRYVLELLGLPLG
DDYAAKRLNGLSGVRNILWSVGGGGASALVGGLTREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVY
EVALALVAQAFIGKKPHLLQDADKQFQQLQQAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDE
CRMWLGLDSEDSQYRNPAIVEFVLENSNRDDNDDLPGLCKLLETWLAGVVFPRFRDTKDKKFKLGDYYDD
PMVLSYLERVEVVQGSPLAAAAAMARIGAEHVKASAMQALQKVFPSRYTDRNSAEPKDVQETVFSVDPVG
NNVGRDGEPGVFIAEAVRPSENFETNDYAIRAGVSESSVDETTVEMSVADMLKEASVKILAAGVAIGLIS
LFSQKYFLKSSSSFQRKDMVSSMESDVATIGSVRADDSEALPRMDARTAENIVSKWQKIKSLAFGPDHRI
EMLPEVLDGRMLKIWTDRAAETAQLGLVYDYTLLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPEN
NATDVRTYTTRYEVFWSKSGWKITEGSVLAS

>gi|20259550|gb|AY091075.1| Arabidopsis thaliana unknown protein (At5g42480) mRNA, complete cds

GATTTAACTTATACTACTCAAAATCAAAATTCCATAAACCCTAGACGACCAAACAGTCTCTTCAATATGT
AAAACAGAACAAAGTTTTTGTAGTAGCCTAAAAAGACACTCCCATGGAAGCTCTGAGTCACGTCGGCATT
GGTCTCTCCCCATTCCAATTATGCCGATTACCACCGGCGACGACAAAGCTCCGACGTAGCCACAACACCT
CTACAACTATCTGCTCCGCCAGCAAATGGGCCGACCGTCTTCTCTCCGACTTCAATTTCACCTCCGATTC
CTCCTCCTCCTCCTTCGCCACCGCCACCACCACCGCCACTCTCGTCTCTCCGCCACCATCTATTGATCGT
CCCGAACGCCACGTCCCCATCCCCATTGATTTCTACCAGGTATTAGGAGCTCAAACACATTTCTTAACCG
ATGGAATCAGAAGAGCATTCGAAGCTAGGGTTTCGAAACCGCCGCAATTCGGTTTCAGCGACGACGCTTT
AATCAGCCGGAGACAGATTCTTCAAGCTGCTTGCGAAACTCTGTCTAATCCTCGGTCTAGAAGAGAGTAC
AATGAAGGTCTTCTTGATGATGAAGAAGCTACAGTCATCACTGATGTTCCTTGGGATAAGGTTCCTGGTG
CTCTCTGTGTATTGCAAGAAGGTGGTGAGACTGAGATAGTTCTTCGGGTTGGTGAGGCTCTGCTTAAGGA
GAGGTTGCCTAAGTCGTTTAAGCAAGATGTGGTTTTAGTTATGGCGCTTGCGTTTCTCGATGTCTCGAGG
GATGCTATGGCATTGGATCCACCTGATTTTATAACTGGTTATGAGTTTGTTGAGGAAGCTTTGAAGCTTT
TACAGGAGGAAGGAGCAAGTAGCCTTGCACCGGATTTACGTGCACAAATTGATGAGACTTTGGAAGAGAT
CACTCCGCGTTATGTCTTGGAGCTACTTGGCTTACCGCTTGGTGATGATTACGCTGCGAAAAGACTAAAT
GGTTTAAGCGGTGTGCGGAATATTTTGTGGTCTGTTGGAGGAGGTGGAGCATCAGCTCTTGTTGGGGGTT
TGACCCGTGAGAAGTTTATGAATGAGGCGTTTTTACGAATGACAGCTGCTGAGCAGGTTGATCTTTTTGT
AGCTACCCCAAGCAATATTCCAGCAGAGTCATTTGAAGTTTACGAAGTTGCACTTGCTCTTGTGGCTCAA
GCTTTTATTGGTAAGAAGCCACACCTTTTACAGGATGCTGATAAGCAATTCCAGCAACTTCAGCAGGCTA
AGGTAATGGCTATGGAGATTCCTGCGATGTTGTATGATACACGGAATAATTGGGAGATAGACTTCGGTCT
AGAAAGGGGACTCTGTGCACTGCTTATAGGCAAAGTTGATGAATGCCGTATGTGGTTGGGCTTAGACAGT
GAGGATTCACAATATAGGAATCCAGCTATTGTGGAGTTTGTTTTGGAGAATTCAAATCGTGATGACAATG
ATGATCTCCCTGGACTATGCAAATTGTTGGAAACCTGGTTGGCAGGGGTTGTCTTTCCTAGGTTCAGAGA
CACCAAAGATAAAAAATTTAAACTCGGGGACTACTATGATGATCCTATGGTTTTGAGTTACTTGGAAAGA
GTGGAGGTAGTTCAGGGTTCTCCTTTAGCTGCTGCTGCAGCTATGGCAAGGATTGGAGCCGAGCATGTGA
AAGCTAGTGCTATGCAGGCACTGCAGAAAGTTTTTCCTTCCCGCTATACAGATAGAAACTCGGCTGAACC
CAAGGATGTGCAAGAGACAGTGTTTAGTGTAGATCCTGTTGGTAACAATGTAGGCCGTGATGGTGAGCCT
GGTGTCTTTATTGCAGAAGCTGTAAGACCCTCTGAAAACTTTGAAACTAATGATTATGCAATTCGAGCTG
GGGTCTCAGAGAGTAGCGTTGATGAAACTACTGTTGAAATGTCCGTTGCTGATATGTTAAAGGAGGCAAG
TGTGAAGATCCTAGCTGCTGGTGTGGCAATTGGACTGATTTCACTGTTCAGCCAGAAGTATTTTCTTAAA
AGCAGCTCATCTTTTCAACGCAAGGATATGGTTTCTTCTATGGAATCTGATGTCGCTACCATAGGGTCAG
TCAGAGCTGACGATTCAGAAGCACTTCCCAGAATGGATGCTAGGACTGCAGAGAATATAGTATCCAAGTG
GCAGAAGATTAAGTCTCTGGCTTTTGGGCCTGATCACCGCATAGAAATGTTACCAGAGGTTTTGGATGGG
CGAATGCTGAAGATTTGGACTGACAGAGCAGCTGAAACTGCGCAGCTTGGGTTGGTTTATGATTATACAC
TGTTGAAACTATCTGTTGACAGTGTGACAGTCTCAGCAGATGGAACCCGTGCTCTGGTGGAAGCAACTCT
GGAGGAGTCTGCTTGTCTATCTGATTTGGTTCATCCAGAAAACAATGCTACTGATGTCAGAACCTACACA
ACAAGATACGAAGTTTTCTGGTCCAAGTCAGGGTGGAAAATCACTGAAGGCTCTGTTCTTGCATCATAAT
ATACTCATATGTAGCATGTCTGAGCTTGCGAGATTCTCTTTGTTTTGTAAATTCTCTCTCTAAGTTAGTG
TTTATAAATGAACACAAAAAAATTAACGTTCAAAAAAAAAAAAAAAA

```
LOCUS       AAM13895                 801 aa            linear   PLN 21-APR-
2002
DEFINITION  unknown protein [Arabidopsis thaliana].
ACCESSION   AAM13895
VERSION     AAM13895.1  GI:20259551
DBSOURCE    accession AY091075.1
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (residues 1 to 801)
  AUTHORS   Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C.,
Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M.,
Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 801)
  AUTHORS   Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C.,
Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M.,
Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Plant Gene Expression Center, 800
Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M.,
Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out
the
            sequencing and annotation of the RAFL cDNAs: Yamada,K.,
Banh,J.,
            Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
            Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
```

```
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed
equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A.
(SSP/PGEC)
            contributed equally to this work as PIs.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..801
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL09-76-G11 (R19395)"
                     /note="This clone is in a modified pBluescript vector
                     (FLC-1) as a BamHI/XhoI insert.
                     ecotype: Columbia"
     Protein         1..801
                     /product="unknown protein"
     CDS             1..801
                     /gene="At5g42480"
                     /coded_by="AY091075.1:114..2519"
ORIGIN
        1 mealshvgig lspfqlcrlp pattklrrsh ntstticsas kwadrllsdf nftsdsssss
       61 fatatttatl vspppsidrp erhvpipidf yqvlgaqthf ltdgirrafe arvskppqfg
      121 fsddalisrr qilqaacetl snprsrreyn eglldeeat vitdvpwdkv pgalcvlqeg
      181 geteivlrvg eallkerlpk sfkqdvvlvm alafldvsrd amaldppdfi tgyefveeal
      241 kllqeegass lapdlraqid etleeitpry vlellglplg ddyaakrlng lsgvrnilws
      301 vggggasalv ggltrekfmn eaflrmtaae qvdlfvatps nipaesfevy evalalvaqa
      361 figkkphllq dadkqfqqlq qakvmameip amlydtrnnw eidfglergl calligkvde
      421 crmwlgldse dsqyrnpaiv efvlensnrd dnddlpglck lletwlagvv fprfrdtkdk
      481 kfklgdyydd pmvlsylerv evvqgsplaa aaamarigae hvkasamqal qkvfpsrytd
      541 rnsaepkdvq etvfsvdpvg nnvgrdgepg vfiaeavrps enfetndyai ragvsessvd
      601 ettvemsvad mlkeasvkil aagvaiglis lfsqkyflks sssfqrkdmv ssmesdvati
      661 gsvraddsea lprmdartae nivskwqkik slafgpdhri emlpevldgr mlkiwtdraa
      721 etaqlglvyd ytllklsvds vtvsadgtra lveatleesa clsdlvhpen natdvrtytt
      781 ryevfwsksg wkitegsvla s
//
```

dbEST Id: 3126415
EST name: 701545606
GenBank Acc: AI998415
GenBank gi: 5845320

CLONE INFO
Clone Id: 701545606
Source: Genome Systems, Inc., a wholly owned subsidiary of Incyte
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

ATAAACACTAACTTAGAGAGAGAATTTACAAAACAAAGAGAATCTCGCAAGCTCAGACAT

GCTACATATGAGTATATTATGATGCAAGAACAGAGCCTTCAGTGATTTTCCACCCTGACT

TGGACCNGAAAACTTCGTATCTTGTTGTGTAGGTTCTGACATCAGTAGCATTGTTTTCTG

GATGAACCAAATCAGATAGACAAGCAGACTCCTCCAGAGTTGCTTCCACCAGAGCACGGG

TTCCATCTGCTGAGACTGTCACACTGTCAACAGATAGTTTCAACAGTGTATAATCATAAA

CCAACCCAAGCTGCGCAGTTTCAGCTGCTCTGTCAGTCCAAATCTTCAGCATTCGCCCAT

CCAAAACCTCTGGTAACATTTCTATGCGGTGATCAGGCCCAAAAGCCAGAGACTTAATCT

TCTGCCACTTGGATACTATATTCTCTGCAGTCCTAGCATCCATTCTGGGAAGTGCTTCTG

AATCGTCAGCTCTGACTGACCCTATGGTAGCGACATCAGNTTCCATAGAAGAAACCATAT
NCTTGCGTTGAAAAGATGAGC

Entry Created: Sep 7 1999
Last Updated: Sep 8 1999

LIBRARY
Lib Name: A. thaliana, Columbia Col-0, rosette-2
Organism: Arabidopsis thaliana
Cultivar: Columbia Col-0
Tissue type: rosette
Develop. stage: 4 - 7 weeks
Vector: pSPORT
R. Site 1: NotI
R. Site 2: SalI
Description: cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted,

ligated to SalI adaptors, digested with NotI, size-selected,
and cloned into the NotI and SalI sites of the pSPORT vector.

SUBMITTER
Name: David Smoller, Ph.D.
Institution: Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.
Address: 4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
E-mail: service@genomesystems.com

CITATIONS
Title: Arabidopsis thaliana Gene Expression MicroArray
Authors: Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle
,M., Brzoska,P., Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton
,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki
,G., Argentine,C., Shah,S., Nobriga,A., Murry,L., Turner,C.,
Krikorian,S., Elder,L., Hanson,D.
Year: 1999
Status: Unpublished

dbEST Id: 5659606
EST name: MtBC10F12F1
GenBank Acc: AL382914
GenBank gi: 9682665

CLONE INFO
Clone Id: MtBC10F12 (T3)
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

CTGGTGTAGCAATTGGACTCATAACTTTAGCTGGTTTGAAGATTTTACCTTCTAAAAATG

GCTCGCCCGTTCTTCACAAAGTGACTGGTTCAGCAATTGCGTCAGATACTATCAATTTAG

GTCCTGTAGGAGATGAAGAATTAGGAGAGCAACTACCAAAAATGAGTGCAATGGTTGCAG

AAGCTCTAGTCCGCAAGTGGCAATATATCACATCCCAAGCTTTTGGACCTGACCATTGCC
TAGGAAGATTGCAAGAGGTGTTGGACGGCCAAATGTTGAAGATATGGACTGATCG

Entry Created: Aug 3 2000
Last Updated: Aug 3 2000

COMMENTS
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).

LIBRARY
Lib Name: MtBC
Organism: Medicago truncatula
Cultivar: Jemalong
Tissue type: arbuscular mycorrhiza
Develop. stage: harvested 3 weeks post inoculation with Glomus intraradices
Vector: pBluescript pSK
R. Site 1: EcoRI
R. Site 2: XhoI
Description: M. truncatula sterilised seeds were germinated for 72h at 25
C, before transplanting into a 1/3 Epoisses soil : 2/3 calcined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LPA8). The plants were watered every day and twice a week with a modified nutrient Long Ashton solution without phosphate but
with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from

polyA+ enriched RNA. The cDNA was directionally ligated into

Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequençage (Genoscope, Evry, France). Note : EST may be of fungal origin.

SUBMITTER

Name:	Genoscope
Institution:	Genoscope - Centre National de Sequencage
Address:	BP 191 91006 EVRY cedex - France
E-mail:	seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

CITATIONS

Title:	Medicago truncatula ESTs from endomycorrhizal roots
Authors:	Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon ,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V., Gamas,P.
Year:	2000
Status:	Unpublished

dbEST Id: 5659607
EST name: MtBC10F12R1
GenBank Acc: AL382915
GenBank gi: 9682666

CLONE INFO
Clone Id: MtBC10F12 (T7)
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

CCCAAGCTTTTGGACCTGACCATTGCCTAGGAAGATTGCAAGAGGTGTTGGACGGCGAAA

TGTTGAAGATATGGACTGATCGAGCAGCTGAGATTGCAGAGCTTGGTTGGTCATATGACT

ACAACTTGGAGGATCTCAACATCGACAGTGTGACCATATCACAGAATGGGCGGCGTGCAG

TAGTGGAAACAACTCTCAAAGAGTCTACCCACCTCACTGCTGTTGGTCATCCACAGCATG

CTACTTCCAACAGCAGAACCTACACAACAAGATATGAAATGTCTTTTTCAGATTCAGGGT

GGAAAATTATTGAAGGAGCTGTCCTTGAGTCGTAATTAGGTTTTGTAATATGTAATATAT

GTCAGGTTAGTACACTTCAATATTAACCCCCTCGAGCCTATGCCCACTGTCTTGTATGTA

CCTGTTGTTTTGTGCATTTTTCAAGCATTTATGTAGTCAGGCTGTAAATACTTGGAGGGT
ATTTGATCAAATAATTATCCGGTTAAAAAAAAAAAAAAAAAAAAAAA

Entry Created: Aug 3 2000
Last Updated: Aug 3 2000

COMMENTS
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).

LIBRARY
Lib Name: MtBC
Organism: Medicago truncatula
Cultivar: Jemalong
Tissue type: arbuscular mycorrhiza
Develop. stage: harvested 3 weeks post inoculation with Glomus intraradices
Vector: pBluescript pSK
R. Site 1: EcoRI
R. Site 2: XhoI
Description: M. truncatula sterilised seeds were germinated for 72h at 25
C, before transplanting into a 1/3 Epoisses soil : 2/3 calcined Terragreen mix in the presence of onion root

fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LPA8). The plants were watered every day and twice a week with a modified nutrient Long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France). Note : EST may be of fungal origin.

SUBMITTER

Name:	Genoscope
Institution:	Genoscope - Centre National de Sequencage
Address:	BP 191 91006 EVRY cedex - France
E-mail:	seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

CITATIONS

Title:	Medicago truncatula ESTs from endomycorrhizal roots
Authors:	Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon ,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V., Gamas,P.
Year:	2000
Status:	Unpublished

dbEST Id: 9071332
EST name: NF119C11IN1F1086
GenBank Acc: BI268376
GenBank gi: 14874230

CLONE INFO
Clone Id: NF119C11IN (5')
Insert length: 660
Plate: 119 Row: C Column: 11
DNA type: cDNA

PRIMERS
Sequencing: TCACACAGGAAACAGCTATGAC
PolyA Tail: Unknown

SEQUENCE

CACGCTTCTCCAAAAAACCTAACCGTCTCCATTCCTCCGCCGTCTCCGCCACCAGTAAAT

GGGCGGAGCGACTCATTTCCGATTTCCAATTCCTCGGCGACACCTCCTCTTCCTCCTCCA

CCACCACCTCCGCCACAGTCACTCTCACTCCTTCTTACCCTCCTCCGATAGAACGCCACG

TGTCACTCCCTCTCGACCTGTACAAAATCCTCGGCGCCGAAACGCATTTTCTCGGTGATG

GTATTCGGAGAGCTTATGAAGCGAAATTCTCGAAGCCTCCTCAGTATGCTTTCAGTAATG

AAGCTTTGATTAGTCGTCGTCAGATTCTTCAAGCTGCTTGTGAAACCCTAGCTGATCCTG

CTTCTAGAAGAGAGTATAATCAAAGCCTCGTCGACGATGAAGACGAAGATGAGGAATCTT

CCATTCTCACTGAAATCCCTTTCGACAAAGTTCCTGGAGCTCTGTGCGTGTTGCAAGAAG

CTGGAGAGACGGAGTTGGTGCTTCGGATTGGAGGGGGTTTACTGAGAGAGAGGTTACCGA

AGATGTTTAAGCAAGATGTTGTGTTGGCTATGGCGCTTGCATATGTTGACGTTTCTAGGG

ATGCTATGGCTTTGTCCCCGCCAGATTTCATTGTTGCTTGTGAGATGCTGGAAAGGGCAT

Entry Created: Jul 18 2001
Last Updated: Jul 18 2001

LIBRARY
Lib Name: Insect herbivory
Organism: Medicago truncatula
Tissue type: local and systemic leaves
Develop. stage: mature
Vector: Lambda Zap
Description: Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled.

SUBMITTER

Name: Korth K
Lab: Dept. of Plant Pathology
Institution: University of Arkansas
Address: 217 Plant Science Building, Fayetteville, AR 72701, USA
Tel: 501 575 5191
Fax: 501 575 7601
E-mail: kkorth@comp.uark.edu

CITATIONS
Title: Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula insect herbivory library
Authors: Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D.
Year: 2000
Status: Unpublished

dbEST Id: 3883556
EST name: si29e11.y1
GenBank Acc: AW472683
GenBank gi: 7042789

CLONE INFO
Clone Id: GENOME SYSTEMS CLONE ID: Gm-r1030-357 (5')
Source: ResGen, Invitrogen Corp.
Insert length: 609
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

AGCGTTGTGTGTGTTGCAGGAAGCTGGAGAGACGGAGCTTGTGCTTGAGATTGGGCAGGG

TTTGCTTAGGGAGAGGTTGCCGAAGACGTTTAAGCAGGATGTTGTGTTGGCTATGGCACT

CGCATTTGTTGACGTGTCAAGGGATGCTTGGCTTGTTCACCGGATTTCATTGCGGCTGTG
AGATGCT

Entry Created: Feb 23 2000
Last Updated: Dec 3 2001

COMMENTS
This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com

LIBRARY
Lib Name: Gm-r1030
Organism: Glycine max
Lab host: DH10B
Vector: pSPORT1
R. Site 1: SalI
R. Site 2: NotI
Description: This cDNA library was constructed from mRNA isolated from immature cotyledons of greenhouse grown plants (individual seed fresh weight of 100-300mg). The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site.
SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna. Note that Gm-r1030 is a re-rack of Gm-c1007.

SUBMITTER
Name: Shoemaker R/Public Soybean EST Project
Lab: Public Soybean EST Project
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS
Title: Public Soybean EST Project
Authors: Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin ,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising ,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R.
Year: 1999
Status: Unpublished

dbEST Id: 5570813
EST name: EST416888
GenBank Acc: BE472035
GenBank gi: 9562526

CLONE INFO
Clone Id: cSTA31L21
Source: Cornell University
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GGAAAGCTTCCTTAACAATGGAGGCATTAACACAGCTAAGCTTTGGCATTTGTACTCCAC

GCCTTTCATCACCATTTCAACTAGCCGCCGCCGGTGGTAAGAAGCCGCCGAGACTCAATG

CCGTTAACGGAGGAGCTAGTAGTGTTACCGGTGGAACAAGTAGTTTACCTACTAACTTCT

CCGCTAGTAAATGGGCGGATCGTCTTCTCGCCGATTTCCAATTCCTTCCTTCCACCACCA

CCTCCGACTCATCGGATTTCCAGAATTCAACTTCTACAACCTCCGTTACGACTATTCCTC

CTCCTGTTGCTCCTTCAGACCACCACATTTCAATGCCTATAGACTTTTATAGAGTGCTTG

GTGCTGAAGCTCACTTCCTCGGTGACGGTATTAGGAGATGCTACGATGCTAGAATTACAA

AGCCTCCGCAGTACGGATACAGTCAGGAAGCATTGATTGGCCGACGGCAGATTCTTCAAG

CTGCTTGTGAAACCCTTGCTGACTCTACCTCTCGTAGAGAGTACAATCAAGGCCTCGCTC

AGCATGAGTTCGATACTATTCTAACTCCTGTCCCCTGGGATAAAGTTCCGGGAGCAATGT
GTGTTTTG

Entry Created: Jul 28 2000
Last Updated: Jul 28 2000

COMMENTS
5 prime sequence

LIBRARY
Lib Name: potato stolon, Cornell University
Organism: Solanum tuberosum
Cultivar: Bintje
Tissue type: axillary buds of stem explants, swelling stolons
Develop. stage: 1 to 3 days
Lab host: SOLR
Vector: pBlueScript SK(-)
R. Site 1: EcoR1
R. Site 2: Xho1
Description: RNA was supplied by Christian Bachem & Beatrix Horvath(Laboratory of Plant Breeding, Dept. of Plant

Sciences, Wageningen University, The Netherlands). Total RNA was isolated from developing axillary buds of potato nodal stem cuttings cultured on medium for the introduction of tuber formation as described in Bachem et al. (Plant Journal 1996). Tissue samples were taken of stages corresponding to growing stolons and the early stages of tuber formation.

SUBMITTER
Name: Research Genetics, Libraries Division
Tel: 1-800-711-6195
E-mail: cdna@resgen.com

CITATIONS
Title: Generation of ESTs from potato swelling stolons
Authors: van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R., Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D., Baker,B.
Year: 1999
Status: Unpublished

dbEST Id: 8892494
EST name: F013P64Y
GenBank Acc: BI120337
GenBank gi: 18004312

CLONE INFO
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GAAGATTTCATGAATGAGGCCTTCTTACGTATGACAGCAGCTGAGCAGGTTGATCTGTTC

GTCACCACGCCAAGTAATATCCCGGCTCAAAATTTTGAAGTTTATGGAGTGGCACTTGCC

CTTGTTGCCCAAGCTTTCATTGGTAAAAAGCCTCATCTCATCACAGATGCTGATAACCTA

TTCGGACAGCTTCAGCAGATTAAGGTAACAAATCAAGGGAGTCTTGTTCCTGTCTTTGGT

TCCATGGAAAACCGTGATATTGACTTTGGGTTGGAGAGGGGCTTTGTTCACTGCTTGTAG
GCCAGCT

Entry Created: Dec 31 2001
Last Updated: Dec 31 2001

LIBRARY
Lib Name: Populus flower cDNA library
Organism: Populus balsamifera subsp. trichocarpa
Organ: flower

SUBMITTER
Name: Erlandsson R
Lab: Department of Biotechnology
Institution: Royal Institute of Technology
Address: Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
E-mail: rikerl@biochem.kth.se

CITATIONS
Title: Gene expression in Populus
Authors: Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao ,R., Jansson,S., Nilsson,O., Sundberg,B., Nilsson,P., Uhlen ,M., Sandberg,G., Lundeberg,J.
Year: 2001
Status: Unpublished

dbEST Id: 1782844
EST name: L30-504T3
GenBank Acc: AI043508
GenBank gi: 3290291

CLONE INFO
Clone Id: L30-504 (5')
Id as DNA: L30-6A504
Id in host: L30-6A504
Insert length: 434
Plate: L30-6 Row: A Column: 12
DNA type: cDNA

PRIMERS
PCR forward: T7
PCR backward: T3
Sequencing: T3
PolyA Tail: Unknown

SEQUENCE

GGGAAACGTGCCTTGGTGGAAGCAACTCTTCAAGAATCAGCGCAGTTAACTGACGTTAAC

CAACCTGAGCATAACGATTCTTACAGCAGAACATACACAACAAGGTACGAGATGTTTCAC

TCCAATGCTGGGTGGAAGATCATAGAGGGAGCTGTCCTCCAATCTTAAGCTGCTGGAAAT

CCAGTCTTGAATGTACATATTTTCACATCATCTGCACATTATGAATGAAGGATGGTATGT

GTTTTCTGGACAGTGGTATTTGATCATGTTGTGTTTATTTTGGTAACAAGTTTTGATCAT

TATCAAAAAGATCACTCTTGTAAGTTAGTTTTTTCCACAATAAATCAACTATTTATATGA
AAGTTTTTATATCAGGACTACTTGCCTTTACTTATATAAACTTTGAGAAATTTTTT
Quality: High quality sequence stops at base: 350

Entry Created: Jul 6 1998
Last Updated: Feb 20 2001
COMMENTS
Poly(A) tail, 18 nt: 417..434
LIBRARY
Lib Name: Ice plant Lambda Uni-Zap XR expression library, 30 hours NaCl treatment
Organism: Mesembryanthemum crystallinum
Tissue type: Leaf, 30 h 0.4M NaCl
Develop. stage: Six week old
Vector: Lambda Uni-Zap XR, Bluescript SK-
R. Site 1: EcoRI
R. Site 2: XhoI
SUBMITTER
Name: Cushman JC
Lab: Department of Biochemistry
Institution: University of Nevada
Address: MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650

E-mail: jcushman@unr.edu

CITATIONS
Title: An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
Authors: Cushman, J.C.
Year: 1997
Status: Unpublished

dbEST Id: 4982897
EST name: AU095068
GenBank Acc: AU095068
GenBank gi: 8857750

CLONE INFO
Clone Id: E51113
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

TGGTGCTTCTCATTTGGGCTGCTGCTGCTGCTATTGCAAAACTTGGTGCTCAAGCTACAG

CTGCACTTGGTACTGTGAAATCAAATGCTATTCAAGCGTTCAACAAGGTTTTNCCATTGA

TAGAACAGTTAGACAGGTCAGCCATGGAAAATACTAAAGATGGCCCTGGGGGATATCTTG

AAAATTTTGACCAGGAAAATGCACCTGCTCATGATTCGAGAAATGCCGCCTTGAAGATTA

TCTCTCTGGCGCACTGTTTGCACTGTTGGCAGTAATTGGGGCCAAATATTTGCCTCGTAA

GAGGCCCCTTTCTGCTATTAGGAGTGAGCATGGATCTGTGGCAGTTGCTAATAGTGTCGA

CTCTACTGATGATCCTGCACTAGATGAAGATCCAGTACATATTCCTAGAATGGATGCGAA
GCTGGCAGAAGATATTGTTCGCAAGTGGCAGAGTATCAAATCTAA

Entry Created: Jun 30 2000
Last Updated: Apr 3 2002

COMMENTS
PROJECT ='RGP'

LIBRARY
Lib Name: Rice immature leaf including apical meristem (under long day
condition)
Organism: Oryza sativa (japonica cultivar-group)
Cultivar: Nipponbare
Develop. stage: immature leaf including apical meristem (under long day condition)

SUBMITTER
Name: Takuji Sasaki
Institution: National Institute of Agrobiological Resources
Address: Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
E-mail: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/

CITATIONS

Title: Rice cDNA from immature leaf including apical meristem
(2000
)
Authors: Sasaki,T., Yamamoto,K.
Year: 2000
Status: Unpublished

dbEST Id: 8592489
EST name: AU183658
GenBank Acc: AU183658
GenBank gi: 14189015

CLONE INFO
Clone Id: E51136
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

ATCATAAGAAGCGCCAAGAAGGGCTTCAAGGTGCGAGAAACATTTTGTGGAGCGTTGGCA

GAGGAGGTATTGCTACCGTTGGAGGAGGATTTTCTCGTGAAGCCTTCATGAACGAGGCTT

TTTTGAGGATGACATCAATTGAACAGATGGATTTCTTTTCAAAAACACCGAATAGCATTC

CTCCTGAATGGTTTGAAATTTACAATGTAGCACTTGCACATGTCGCTCAAGCAATTATAA

GTAAAAGGCCACAATTCATCATGATGGCGGATGATCTTTTTGAACAACTCCAGAAGTTCC
ACATAGGTC

Entry Created: May 22 2001
Last Updated: Apr 3 2002

COMMENTS
PROJECT = 'RGP'

LIBRARY
Lib Name: Rice immature leaf including apical meristem (under long day
condition)
Organism: Oryza sativa (japonica cultivar-group)
Cultivar: Nipponbare
Develop. stage: immature leaf including apical meristem (under long day condition)

SUBMITTER
Name: Takuji Sasaki
Institution: National Institute of Agrobiological Resources
Address: Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
E-mail: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/

CITATIONS
Title: Rice cDNA from immature leaf including apical meristem (2001
)
Authors: Sasaki,T., Yamamoto,K.
Year: 2001

FIG. 8 continued 28/110

Status: Unpublished

dbEST Id: 2462373
EST name: AU058418
GenBank Acc: AU058418
GenBank gi: 4714451

CLONE INFO
Clone Id: E51113_1A
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

ATCATAAGAAGCGCCAAGAAGGGCTTCAAGGTGCGAGAAACATTTTGTGGAGCGTTGGCA

GAGGAGGTATTGCTACCGTTGGAGGAGGATTTTCTCGTGAAGCCTTCATGAACGAGGCTT

TTTTGAGGATGACATCAATTGAACAGATGGATTTCTTTTCAAAAACACCGAATAGCATTC

CTCCTGAATGGTTTGAAATTTACAATGTAGCACTTGCACATGTCGCTCAAGCAATTATAA

GTAAAAGGCCACAATTCATCATGATGGCGGATGATCTTTTTGAACAACTCCAGAAGTTCA
ACATAGGTTCTCATTATGCTTATGATAATGAGATGG

Entry Created: Apr 29 1999
Last Updated: Apr 1 2002

COMMENTS
PROJECT ='RGP'

LIBRARY
Lib Name: Oryza sativa Nipponbare immature leaf including apical meristem (under long day condition)
Organism: Oryza sativa (japonica cultivar-group)
Cultivar: Nipponbare
Develop. stage: immature leaf including apical meristem (under long day condition)

SUBMITTER
Name: Takuji Sasaki
Institution: National Institute of Agrobiological Resources
Address: Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
E-mail: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/

CITATIONS
Title: Rice cDNA from immature leaf including apical meristem
Authors: Sasaki,T., Yamamoto,K.
Year: 1997
Status: Unpublished

dbEST Id: 5613895
EST name: WHE0365_C09_F17ZS
GenBank Acc: BE490117
GenBank gi: 9609650

CLONE INFO
Clone Id: WHE0365_C09_F17
DNA type: cDNA

PRIMERS
Sequencing: Stratagene SK primer
PolyA Tail: Unknown

SEQUENCE

CAGTGCTTGCAATTGGAGGGCACTTACTGGAGGACCGCCCGCCCAAGCGGTTCAAGCAGG

ATGTGGTGCTGGCAATGGCGCTCGCTTATGTGGATCTATCAAGGGACGCAATGGCGGCTA

GCCCTCCAGATGTAATCCGCTGCTGTGAGGTGCTTGAAAGGGCTCTCAAGCTTTTGCAGG

AGGATGGGGCAATCAATCTCGCACCTGGTTTGCTCTCACAAATTGATGAAACTCTGGAGG

ATATCACACCTCGTTGTGTTTTGGAGCTTCTTGCCCTTCCTCTTGATGAAAAACATCAGA

ATGAACACCAAGAAGGTCTTCGTGGTGTGAGAAACATTTTGTGGAGTGTTGGCAGAGGAG

GTATTGGTACTGTTGGAGGAGGATTTTCGCGTGAAGCCTACATGAATGAAGCCTTCCTGC

AGATGACATCGGCGGAGCAGATGGATTTCTTCTCAAAAACACCGAATAGCATACCGCCTG
AATGGTTTGAAATCTATAGCGTGGCACTTGCAAATGTTGCTCAAGCAATTGTAAGTA

Entry Created: Jul 31 2000
Last Updated: Jul 31 2000

COMMENTS
Sequence have been trimmed to remove vector sequence and
low
quality sequence with phred score less than 20

LIBRARY
Lib Name: Wheat cold-stressed seedling cDNA library
Organism: Triticum aestivum
Cultivar: Chinese Spring
Tissue type: Seedling
Develop. stage: Five-day old seedling
Lab host: E. coli SOLR
Vector: Lambda Uni-ZAP XR, excised phagemid
R. Site 1: EcoRI
R. Site 2: XhoI
Description: Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were

transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).

SUBMITTER
Name: Olin Anderson
Institution: US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
Address: 800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
E-mail: oandersn@pw.usda.gov

CITATIONS
Title: The structure and function of the expressed portion of the wheat genomes - Cold-stressed seedling cDNA library
Authors: Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch ,C.J., Seaton,C.L., Tong,J.C.
Year: 2000
Status: Unpublished

dbEST Id: 8348091
EST name: WHE2493_E05_J09ZS
GenBank Acc: BG607272
GenBank gi: 13657255

CLONE INFO
Clone Id: WHE2493_E05_J09
DNA type: cDNA

PRIMERS
Sequencing: Stratagene SK primer
PolyA Tail: Unknown

SEQUENCE

ACACCTCGTTGTGTTTTGGAGCTTCTTGCCCTTCCTCTTGATGAAAAGCACCAGAGTAAA

CGCCAAGAAGGTCTTCGTGGTGTGAGAAACATTTTGTGGAGTGTTGGTAGAGGAGGTATT

GCTACTGTTGGAGGAGGATTTTCNCGTGAAGCCTACATGAATGAGGCCTTTTTGCAGATG

ACATCAGCGGAGCAGATGGATTTCTTTTCAAAAACGCCAAATAGCATACCACCTGAATGG

TTTGAAATCTATAGTGTGGCACTCGCAAATGTTGCTCAAGCAATTGTAAGTAAAAGGCCA

NAGCTCATCATGGTGGCAGATGATCTTTTCGAACAGCTCCAGAAGTTCAATATAGGTTCT
CAATATGCTTATGATAATGAATTGGATCTTGTGTTGGAAAGGGCACTTTGCTCATTGC

Entry Created: Apr 17 2001
Last Updated: Apr 17 2001

COMMENTS
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20

LIBRARY
Lib Name: Triticum monococcum early reproductive apex cDNA library
Organism: Triticum monococcum
Cultivar: DV92
Tissue type: Early reproductive apex
Develop. stage: Seven week-old plants
Lab host: E. coli XLOLR
Vector: Lambda Uni-ZAP XR, excised phagemid
R. Site 1: EcoRI
R. Site 2: XhoI
Description: The tissue, total RNA, and poly(A) RNA were prepared from apex at double-ridge stage to terminal-spikelet stage during transition from vegetative state to flower state, a cDNA library was made, and the cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova, J. Dubcovsky). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).

SUBMITTER
Name: Olin Anderson
Institution: US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
Address: 800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
E-mail: oandersn@pw.usda.gov

CITATIONS
Title: The structure and function of the expressed portion of the wheat genomes - Early reproductive apex cDNA library from Triticum monococcum
Authors: Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S.
, Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Stamova,B., Tong,J.C.
Year: 2001
Status: Unpublished

dbEST Id: 9919900
EST name: HVSME10017D16f
GenBank Acc: BI949952
GenBank gi: 16291659

CLONE INFO
Clone Id: HVSME10017D16f
Source: CUGI
DNA type: cDNA

PRIMERS
Sequencing: AATTAACCCTCACTAAAGGG
PolyA Tail: Unknown

SEQUENCE

GCGAGCATGAGTCCGTGGCAGTTGCTAATGTTGTTGACTCAGGTGATGATGACGAACCAG

ATGAGCCCATACAGATTCCTAAAATGGATGCGAAGCTGGCAGAAGATATTGTTCGCAAGT

GGCAGAGCATCAAATCCAAGGCCTTGGGATCAGATCATTCTGTTGCATCATTGCAAGAGG

TTCTTGATGGCAACATGCTGAAGGTATGGACGGACCGAGCAGCAGAGATCGAGCGCAAAG

GCTGGTTCTGGGACTACACGCTGTCCAACGTGGCGATCGACAGCATCACCGTCTCCCTGG

ACGGACGGCGGGCGACCGTGGAGGCGACAATTGAGGAGGCGGGTCAGCTCACCGACGCAA

CCGACCCCAGGAACGATGATTTGTACGACACTAAGTACACCACCCGGTACGAGATGGCCT

TCACCGGACCAGGAGGGTGGAAGATAACCGAAGGCGCAGTCCTCAAGTCGTCATAGGGCG
Quality: High quality sequence stops at base: 474

Entry Created: Oct 19 2001
Last Updated: Oct 19 2001

COMMENTS
Total hq bases = 422

LIBRARY
Lib Name: Hordeum vulgare spike EST library HVcDNA0012 (Fusarium infected)
Organism: Hordeum vulgare
Cultivar: Morex
Tissue type: Spike
Lab host: TJC121
Vector: pBluescript SK(-)
R. Site 1: EcoRI
R. Site 2: XhoI
Description: Plants were grown at the University of Minnesota in the GJ Muehlbauer lab; spikes were harvested and snap frozen at 0, 1, 2, 3, 4, 5, 6, and 8 days after Fusarium graminearum inoculation (Heinen). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all eight RNA pools were

combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi, Fenton, Malatrasi). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)

SUBMITTER
Name: Wing RA
Lab: Clemson University Genomics Institute
Institution: Clemson University
Address: 100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
E-mail: rwing@clemson.edu

CITATIONS
Title: Development of a genetically and physically anchored EST resource for barley genomics: Fusarium infected Morex spike cDNA library
Authors: Wing,R., Muehlbauer,G.J., Close,T.J., Kleinhofs,A., Wise,R., Heinen,S., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Fenton,R.D., Malatrasi,M., Choi,D.W., Oates,R., Main,D.
Year: 2001
Status: Unpublished

dbEST Id: 8864363
EST name: AV833644
GenBank Acc: AV833644
GenBank gi: 14525733

CLONE INFO
Clone Id: bags1d11
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GAAACTCTGGNNGNAGATCACCCCTCGTTGTGTTTTAGAGCTTCTTGCCCTTCCTCTTGA

CGAGNAAGCACCAGAGTAAACGCCAAGNAAGGTCTTCGTGGTGTGAGAAACATTTTGTGG

AGTGTTGGTAGAGGAGGTATTGCTACTGTTGGTGGAGGATTTTCACGGGAAGCCTACATG

AATGAGGCCTTTTTGCAGATGACATCAGCTGAGCAGATGGATTTCTTTTCAAAAACGCCG

AATAGCATACCACCTGAATGGTTTGAAATCTATAGCGTGGCACTCGCAAATGTTGCTCAA

GCAATTGTAAGTAAAAGGCCAGAGCTCATCATGGTGGCAGATGATCTTTTCGAACAGCTC

CAGAAGTTCAATATCGGTTCTCAATATGCTTATGGTAACGAGATGGATCTTGCGTTGGAA

AGGGCACTTTGCTCATTGCTTGTGGGAGACATTAGCAACTGCAGAACTTGGCTTGCGATT

GATAATGAATCTTCACCACATAGAGACCCGAAAATTGTAGAGTTTATTGTGAACAACTCT

AGCATTGACCACCAGGAGAATGATCTTCTTCCAGGCCTGTGTAAGCTTTTGGAGACTTGG
CTTGTCTCAGAGGTTTTCCCTA

Entry Created: Jun 22 2001
Last Updated: Jun 22 2001

COMMENTS
Sato,K., Saisho,D., Takeda,K., Shini,T. and Kohara,Y.
Direct
submission;
database:http://www.shigen.nig.ac.jp/barley/Barley.html

LIBRARY
Lib Name: K. Sato unpublished cDNA library: Hordeum vulgare subsp. vulgare shoots germination
Organism: Hordeum vulgare subsp. vulgare
Cultivar: Haruna Nijo
Tissue type: shoots
Develop. stage: germination

SUBMITTER
Name: Kazuhiro Sato
Lab: Research Institute for Bioresources

Institution: Okayama University, Barley Germplasm Center
Address: Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
E-mail: kazsato@rib.okayama-u.ac.jp,
URL:http://www.rib.okayama-u.ac.jp/barley/

CITATIONS
Title: Barley EST sequencing project in NIG and Okayama Univ.
Authors: Sato,K.
Year: 2001
Status: Unpublished

dbEST Id: 10841891
EST name: AV921157
GenBank Acc: AV921157
GenBank gi: 18216936

CLONE INFO
Clone Id: bags1d11 (3')
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

TGGCTTCACCTGNAAATCCAGCACTAAGTTTCTCTTATCACCAACCCAAGGATCTCTTCT

AGCCTAGCAATAATCCGAATAGAACACACCGAAAAACAAAGCTCATCGCTGACTAACTGA

CTAACCAAACTATCTCCGTCTTCCAAACTGACAAGAGCCTAGACTAGACTGCTTATTTAC

ACACCAGAAAAACACGGGAGGAATCAATCAACAAGGTTTACTGCACGCTGAACGCCCTAT

GACGACTTGAGGACTGCGCCTTCGGTTATCTTCCACCCTCCTGGTCCGGTGAAGGCCATC

TCGTACCGGGTGGTGTACTTAGTGTCGTACAAATCATCGTTCCTGGGGTCGGTTGCGTCG

GTGAGCTGACCCGCCTCCTCAATTGTCGCCTCCACGGTCGCCCGCCGTCCGTCCAGGGAG

ACGGTGATGCTGTCGATCGCCACGTTGAACAGCGTGTAGTCCCAGAACCAGCCTTTGCGC

TCAATCTCTGCTGCTCGGTCTGTCCATACCTTCAGNATGTTGCCATCAAGAACCTCTTGC

AATGATGCAACAGAATGATCTGATCCCAAGGCCTTGGATTTGATGCTCTGCCACTTGCGA
ACAA

Entry Created: Jan 18 2002
Last Updated: Jan 18 2002

LIBRARY
Lib Name: K. Sato unpublished cDNA library, cv. Haruna Nijo germination shoots
Organism: Hordeum vulgare subsp. vulgare
Cultivar: Haruna Nijo
Tissue type: shoots
Develop. stage: germination

SUBMITTER
Name: Tadasu Shin-i
Lab: Center For Genetic Resource Information
Institution: National Institute of Genetics
Address: 1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
E-mail: tshini@genes.nig.ac.jp

CITATIONS
Title: Barley EST sequencing project in NIG and Okayama Univ
Authors: Sato,K., Saisho,D., Takeda,K.
Year: 2002
Status: Unpublished

dbEST Id: 6212986
EST name: OV1_8_A03.g1_A002
GenBank Acc: BE917942
GenBank gi: 10420549

CLONE INFO
DNA type: cDNA

PRIMERS
Sequencing: PolyTMix
PolyA Tail: no

SEQUENCE

TATGGGTCTGTGGCAGTTGCTGACTCTGTTGATGGTCTGGGAGCAGATGAAGAGCCACTA

GAAATTCCTAGAATGGATGCAAAGTTGGCTGAAGATATTGTTCGCAAGTGGCAAAGTATC

AAGTCCAAGGCTTTGGGGCCAGAACACACTGTCACGGCATTGCAAGAGATCCTCGATGGC

AACATGCTGAAGGTATGGATGGACCGAGCCACAGAGATTGAGCGTCACGGTTGGTTCTGG

GAATACACACTCTCCGACGTGACGATCGACAGTATCACCGTCTCCATGGACGGTCGACGG

GCAACTGTGGAGGCGACGATTGAGGAGATGGGCCAACTTACCGACGTAGCAGACCCAAAG

AACAACGACGCCTACGACACAAAGTACACCGCTCGGTACGAGATGAGCTACTCCAAGTCC

GGAGGGTGGAGGATCACCGAAGGAGCAGTCCTCAAGTCGTAGAACGGTCGTGCAGCAGGA

GTAGGCGAGTAGGGGTTGCTCAACTCCCATTCTTTTTTCTTTTGCACCAGTGTATGTAAA

TAAACAGTGTGAGCACAGGTTCTTTTCTCTCCTGGAGAGAGTTTGGTTAGGTTGATTAGT
GATGAGTTCCTGAGGCCGAGAGAATTTGTCATCTAGTTTGTATTGATAGAGAT
Quality: High quality sequence starts at base: 17
Quality: High quality sequence stops at base: 640

Entry Created: Sep 29 2000
Last Updated: Sep 29 2000

COMMENTS
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

LIBRARY
Lib Name: Ovary 1 (OV1)
Organism: Sorghum bicolor
Organ: Mix of ovaries of varying immature stages from 8-week-old plants
Vector: pBluescript II from Lambda Zap II
R. Site 1: XhoI
R. Site 2: EcoRI
Description: The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass

excision.

SUBMITTER
Name: Cordonnier-Pratt MM
Lab: Laboratory for Genomics and Bioinformatics
Institution: The University of Georgia, Department of Plant Biology
Address: Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
E-mail: mmpratt@uga.edu

CITATIONS
Title: An EST database from Sorghum: ovaries of varying immature stages
Authors: Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M., Pratt,L.H.
Year: 2000
Status: Unpublished

dbEST Id: 6213567
EST name: OV1_8_A03.b1_A002
GenBank Acc: BE918523
GenBank gi: 10421712
CLONE INFO
DNA type: cDNA
PRIMERS
Sequencing: JEN REV
PolyA Tail: no
SEQUENCE

GCACGAGGATAGAACAGCTAGACAGATCAGGCAAGGATACCCCAGGTGATGATCTTGAGA

AATCTCTTGAAAAACTTGCCCAAGAAATGTTGCTGGAGATGCTATCCATGATTCCAAAAA

TGCCGCTTTGAAGATTATCTCTGCTGGTGCACTGTTTGCACTATTTGCAGTAATAGGTCT

GAAGTGCTTGCCTCGTAAGAAGTCACTTCCTGCTCTTAAGAGCGAATATGGGTCTGTGGC

AGTTGCTGACTCTGTTGATGGTCTGGGAGCAGATGAAGAGCCACTAGAAATTCCTAGAAT

GGATGCAAAGTTGGCTGAAGATATTGTTCGCAAGTGGCAAAGTATCAAGTCCAAGGCTTT

GGGGCCAGAACACACTGTCACGGCATTGCAAGAGATCCTCGATGGCAACATGCTGAAGGT

ATGGATGGACCGAGCCACAGAGATTGAGCGTCACGGTTGGTTCTGGGAATACACACTCTC
CGACGTGACGATCGACAGTATCACCGTCTCCATGGACGGTCGACGGGCAACTGTG
Quality: High quality sequence stops at base: 447
Entry Created: Sep 29 2000
Last Updated: Sep 29 2000.
COMMENTS

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

LIBRARY
Lib Name: Ovary 1 (OV1)
Organism: Sorghum bicolor
Organ: Mix of ovaries of varying immature stages from 8-week-old plants
Vector: pBluescript II from Lambda Zap II
R. Site 1: XhoI
R. Site 2: EcoRI
Description: The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision.

SUBMITTER
Name: Cordonnier-Pratt MM
Lab: Laboratory for Genomics and Bioinformatics
Institution: The University of Georgia, Department of Plant Biology
Address: Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
E-mail: mmpratt@uga.edu

CITATIONS

Title:	An EST database from Sorghum: ovaries of varying immature stages
Authors:	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M., Pratt,L.H.
Year:	2000
Status:	Unpublished

dbEST Id: 11076385
EST name: 952021B01.x1
GenBank Acc: BM498278
GenBank gi: 18649459
CLONE INFO
Plate: 952021 Row: B Column: 01
DNA type: cDNA
PRIMERS
PolyA Tail: Unknown
SEQUENCE

GCCACAGGCCGCCACCGCCTGGCCCCTCCACCTGCCGCTCCGCCAGCCGCTGGGCCGACC

GCCTCTTCGCCGACTTCCACCTCCTCCCCGCCGCCGCCGACCCGCCAGCCGCGGCCTCCT

CTTCCTCCTCGTCCCCGTTCGTCCCGATCTTCCCCGAAGCCGCCGACCGCGCCTTGCCCC

TCCCGGTCGACTTCTACAAGATTCTTGGTGCGGAGCCACATTTCCTAGGCGATGGCATTC

GGAGGGCGTTCGAGTCGCGGATAGCTAAGCCACCTCAGTATGGGTACAGCACAGAAGCTC

TTGCTGGGCGACGGCAAATGCTGCAGATTGCCCATGATACTCTCACAAACCAGAGCTCGC

GCACCGAGTACGACCGTGCGCTTTCCGAGGACCGTGATGCGGCACTCACCATGGATGTTG
CCTGGGATAAGGTTCCAGGTGTGCTGCGTGTGCTTCAGGAGGCTGGGGAGGCACAACTG
Entry Created: Feb 11 2002
Last Updated: Feb 11 2002
LIBRARY
Lib Name: 952 - BMS tissue from Walbot Lab (reduced rRNA)
Organism: Zea mays
Cultivar: BMS (Black Mexican Sweet)
Tissue type: suspension culture
Develop. stage: mixed logarithmic and stationary growth phases
Lab host: DH10B
Vector: pUC19
R. Site 1: EcoRI
R. Site 2: EcoRI
Description: The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive
clones.
SUBMITTER
Name: Walbot V
Lab: Department of Biological Sciences
Institution: Stanford University
Address: 855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
E-mail: walbot@stanford.edu

CITATIONS
Title: Maize ESTs from various cDNA libraries sequenced at Stanford
University.
Authors: Walbot,V.
Year: 1999
Status: Unpublished

dbEST Id: 11076864
EST name: 952021B01.y1
GenBank Acc: BM498757
GenBank gi: 18649938

CLONE INFO
Plate: 952021 Row: B Column: 01
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

AGCAATGTGGGCAAGTGCGACACTATAGATCTCAAACCATTCAGGTGGTATGCTATTCGG

TGTTTTAGAGAAGAAATCCATCTGCTCAGCTGATGTCATCTGCAAGAAAGCCTCATTCAT

GAAGGCCTCACGAGAAAATCCTCCTCCAACAGTAGCAATACCACCCCTGCCAACACTCCA

CAATATGTTTTTTGCACCTTGCAGACCTTCTTGGCGTTTATTTTTATGTTTTTCATCAGT

AGGAAGAGCAAGAAGCTCCAATACACAACGAGGTGTAATCTCCTCCAAAGTTTCATCAAT

CTGTGCAAGCAGTTCAGGTGCAAGATTGCTTGCACCATCCTCCTGCAGGAGCTTCAGTGC

CCTCTCAAGCACCTCACAACAGCAGATTACATCTGGAGGGCTTGCTGCCATAGCATCCCT
TGATATGTCCACATAAGCCAATGCCA

Entry Created: Feb 11 2002
Last Updated: Feb 11 2002

LIBRARY
Lib Name: 952 - BMS tissue from Walbot Lab (reduced rRNA)
Organism: Zea mays
Cultivar: BMS (Black Mexican Sweet)
Tissue type: suspension culture
Develop. stage: mixed logarithmic and stationary growth phases
Lab host: DH10B
Vector: pUC19
R. Site 1: EcoRI
R. Site 2: EcoRI
Description: The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive
clones.

SUBMITTER
Name: Walbot V

Lab: Department of Biological Sciences
Institution: Stanford University
Address: 855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
E-mail: walbot@stanford.edu
CITATIONS
Title: Maize ESTs from various cDNA libraries sequenced at Stanford
University.
Authors: Walbot,V.
Year: 1999
Status: Unpublished

dbEST Id: 3713166
EST name: 707034D03.x3
GenBank Acc: AW331058
GenBank gi: 6827415

CLONE INFO
Plate: 707034 Row: D Column: 03
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

CGCGTCGACGTATAGAGTCTGCATCCATGTTGCCTTGAATGAAGCGTCTGCAAAAGAAGG

CTCTTTTATCACCAGTCGTGTCAGGAAGCATTTTGAAAATATATCAAAATTTCTTTGGCT

GAGTGATAGGCCTAATTCAAATAGCAAAGGAAGTGATAAACACCCAGCGGTTAATGATAT

TACTGCTGCAGTTTGCAAGCAAAAGATGGATATTCAAGAAGCAGAAACACTTGTAAAACA

GTGGCAAGACATAAAATCTGAAGCTCTTGGCCCTGACTATCAAACTGACATGCTACCTGA

GATTCTTGATGGTTCAATGCTCTCTAAGTGGGAAGACTTAGCGTTATTAGCAAAGGACCA

GTCTTGCTATTGGAGATTTGTGCTGCTAAATCTTAATGTTGTTCGAGCCGAGATAATCTT

GGATGAAATAGGTGCTGGTGAGGCAGCAGAAATTGATGCTGTACTTGAGGAAGCGGCTGA

GCTTGTTGACGATTCCCAGCCCAAGAAACCGAGTTATTACAGCACATATGAAGTTCAGTA

CGTATTGAGGAGGCAGAATCATGGATCTTGGAAAATCTCCGAGGCTGCTGTCCGGGACCT
GACGTGATTTCTGCCAACTCGGCAAACGGGCTACACAACCATTGGCGTATAGGCGGC

Entry Created: Jan 31 2000
Last Updated: Jan 31 2000

LIBRARY
Lib Name: 707 - Mixed adult tissues from Walbot lab (SK)
Organism: Zea mays
Cultivar: W23
Organ: tassel, kernel, silk, husk, root, leaf
Tissue type: tassel, kernel, silk, husk, root, leaf
Develop. stage: adult
Lab host: DH10B
Vector: pGAD10
R. Site 1: EcoRI
Description: cDNA library from fully differentiated maize tissues from an
active Mutator plant. Tissue ratio is 4/2/1/1/1/1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned.

SUBMITTER
Name: Walbot V

Lab: Department of Biological Sciences
Institution: Stanford University
Address: 855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
E-mail: walbot@stanford.edu

CITATIONS
Title: Maize ESTs from various cDNA libraries sequenced at Stanford University.
Authors: Walbot,V.
Year: 1999
Status: Unpublished

dbEST Id: 5882137
EST name: Cri2_3_H15_SP6
GenBank Acc: BE641509
GenBank gi: 9959174

CLONE INFO
Clone Id: Cri2_3_H15 (5')
Plate: Cri2_3 Row: H Column: 15
DNA type: cDNA

PRIMERS
Sequencing: SP6
PolyA Tail: Unknown

SEQUENCE

GTGGTGTCTTTGCTCGTGTTCCTGGATACACAAGGGATGAGTATATGAAGGCAGCTTTTT

CTCGAATGACAGCTGCTGAGCAAGTAGCTTTGTTCACAAATACACCCAGTAATATCCCAG

CAGAGAGTTCTGAGGTTTACACAGTTGCGCTTGCTCACATAGCAGAGGGATTTGTTGCAA

AGAAGCCGCAATTGATTCAGGAAGCTGATTCACTCTTTCTTCAGCTTCAGCGAACAAATG

CCTCATCATCTAGTTTGCTAGTTACTGGTGGTCTACGGCCATTATCAAGTCTGCAGCTTG

ATTTTGCTTTTGAACGAGCCATGTGCAAACTGCTCCTAGGAGAACTGGATGGTTGTCGTG

CATGGCTAGGTTTGGATGATACAAACTCTCCATATAGAGACCCTGCAGTGACTGATTTTG

TTATAGCTAATTCTTTTGGAAGTGAGGAAGGTGATTATTTACCAGGCCTTTGCAAGTTGT

TGGAAAGTTGGTTGAGGGAAGCGGTGTTTTTCCCCAACCCGTCAACAGAAAAGTGGAGGT

ACAAGTTGAGGGAGTATTTTTTATGATGCAAGGAGAAAAAAAGCCGCCGTGAATTTTTTC

GCGGGGGGCGCTATGAAAAAATATATTCAACCTTTTTTTGTTGGGGCGTCGTCTACAAAG

AATGATGGAGTGTCATTGTTGCTTTTGAGGTGACGAAGGGGCGGCGCTCCTCTTTAAGGG

ATCGTCCGTGGGGGCGCGCGCTCCCATATCGCCATCTTCGGGACACCTTGTTCGTGGGTC

AAATGGTGATGTCTTTTTTACCACGAACGTCACATTATTCTTATAATATAAGCGTGCGGC
AGCACTCTCAGCTTCGACGAAACAGCCTAAA
Entry Created: Sep 1 2000
Last Updated: Sep 1 2000
LIBRARY
Lib Name: Ceratopteris Spore Library
Organism: Ceratopteris richardii
Cultivar: Brogn
Tissue type: Gametophyte
Cell type: Spore
Develop. stage: 20 hours after germination initiation
Vector: pCMVSPORT6
Description: EST sequence from cDNA library. cDNA library constructed

from mRNA isolated from C. richardii spores that had developed for 20 hours after their germination had been initiated by white light.

SUBMITTER
Name: Roux SJ
Lab: Section of Molecular Cell and Developmental Biology
Institution: University of Texas
Address: Biology Building, Room 16, Austin, TX 78712, USA
Tel: 512 471 4238
Fax: 512 232 3402
E-mail: sroux@uts.cc.utexas.edu

CITATIONS
Title: Expressed sequence tags of cDNA clones from a C. richardii library
Authors: Chatterjee,A., San Miguel,P., Stout,S.C., Banks,J., Roux ,S.J.
Year: 2000
Status: Unpublished

dbEST Id: 9279697
EST name: gc56a02.y1
GenBank Acc: BI437111
GenBank gi: 15261801

CLONE INFO
Clone Id: PEP_SOURCE_ID:PPN190104 (5')
Source: University of Leeds (UK) & Washington University in St. Louis (USA)
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GAGAACGGAAGCTTTAGAAGTGGAGGTTGTCCCCAAAATGGATGCTAGGTTGGCGGAAAT

TATGGTTCGAAGATGGCAAGCAGCTAAAGCTCGAGCACTTGGTTCTGCTCATGATATGGC

GGCTCTTCCTGAGGTGCTGGAGGGCGAGATGCTGAAGAGCTGGACAGACCGTGTTAGTGA

CGTCAAGAGAAATGGTTGGTTTTGGGAATACACTCTCCTTGGTCTTCACATTGATAGTGT

AACAGTAAGTGACGATGGGAGGCGAGCAACTGCGGAAGCCACTTTGCAAGAGGCAGCCCG

CTTGGTGGACCGCAACAACCCTGACCACAATGATTCTTATAGAAGCACTTACACTACGCG

ATATGACCTCCGGCATGGCATAGATGGTTGGCGAATCAATGGAGGAGCTGTGCTGCGTAC

TTGATTCTGAGATTTTCATCTCCGGATCATGTTGACTTGTAGGCAGATCGACTAGTTGCA

ACCCTTGCATGCTACGAATGAGTAGTCTTTTTGGATATTTTGATCCATCATGCAGCTTTG
A

Quality: High quality sequence stops at base: 424

Entry Created: Aug 21 2001
Last Updated: Aug 21 2001

COMMENTS
Libraries were constructed by Dr. Stavros Bashiardes as part
of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

LIBRARY
Lib Name: Moss EST library PPN
Organism: Physcomitrella patens
Tissue type: protonemata: 7 day old tissue auxin treated
Lab host: DH10B
Vector: pBluescript SK-
R. Site 1: EcoRI

R. Site 2: XhoI
Description: Construction of the cDNA library was carried out using Stratagenes 'UniZAP - cDNA synthesis kit'. cDNA was constructed using an oligo dT primer/linker that contains a XhoI site within it. Following ds cDNA synthesis, EcoRI adapters were ligated to the blunt ends and sample was digested with XhoI. The result is cDNA with an EcoRI sticky end on one side and a XhoI sticky end on the other. This cDNA was ligated directionally in UniZAP arms. The vector is designed containing the pBluescript sequence as well as lambda DNA and cDNA is cloned within this pBluescript sequence. The vector was then packaged using Gold gigapackaging extracts. Library was grown in XLIBlue MRF' cells and amplified. The library was excised by mass excision using Stratagens 'Mass excision kit' that uses exassist as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids. SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered by using Quiagen Midi prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation.

SUBMITTER
Name: Ralph Quatrano
Lab: Leeds/Wash U Moss EST Project
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS
Title: Leeds/Wash U Moss EST Project
Authors: Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R., Wilson,R.
Year: 1999
Status: Unpublished

Prochlorococcus marinus sp. MED4 analysis files DRAFT
Produced for the Joint Genome Institute Microbial Sequencing program.
N.B.: These pages subject to frequent change - work in progress.

http://genome.ornl.gov/cgi-bin/JGI_microbial/gene_viewer.cgi?org=pmar_med&chr=1&contig=pmar_med&gene=533

Version 1 - pmar_med Gene 533
Gene Finders
Strand = r
Stop Location = 1236816
Stop Codon = tag

Gene Modeler	Start Location	Start Codon
Generation	1238441	atg
Glimmer	1238837	ttg
Critica	1238924	ttg

MRNA
ttggaacttccattagatcactttcgtttaataggcgtaagcccctcagcaacatctgaggaaatattaagggct
ttcca
attacgcttggataaaactcctgatgaaggattcacgtacgaggttttaactcaaaggtcggaattgcttcgcct
tactg
cagatttgcttacagatccagatagtagaagagattacgaaaatttattactaaatggagcatcaggtttagatt
tatct
tccaatagagaggttgcaggattaattctcctttgggaatcgggctcttctaaagaagcctttaaaataacaaga
aaagc
attgcaaccccccccaaactcctgcattgggtagcagtagagaagctgatcttaccttgttagcggctttaacatc
tagag
atgctgcaatacaagagcaagatcaaagatcttactcaaatgctgcagatttttacaagaaggcatacagcttc
ttcaa
agaatgggcaaactaggggaattacggaaaactcttgaggaggacttagtgtcgcttcttccgtatcgaattctt
gattt
gttaagtagagatctaaatgattatgactcgcataaaaaaggtttaagtatgctggaaaatttaataatcaaaag
aggtg
gattagaaggaaaaaataaatctgaatataatgattttctaaatcagcaagaatttgaatctttctttcaacaaa
taaag
ccattcttgactgttcaggatcagatagatttatttttagaattacaaaaaaggggttcaagtgaagcaggattt
ttagc
ttttttatctttaacagcaattggttttgcaagaagaaaacctgcaaaattattcgaagctcgaaaaatattaaa
aaaac
taaatttatcaggacttgactcaatgccattaataggttgccttgatttgcttttagcagatgttgagcaatcct
cagca
aggtttttaagtagttccgatgagaagttaagagattggttgaataattatcctggagaaaaattagaagcaata
tgtat
tttttgtaaaaattggttagaaaatgatgttttggttggttatagggatattgatttaaaagaaatcgatttaga
ctctt
ggtttgaagatagagaaatccaagaatttattgagcaaatagaaaagaagtcaaatagaactgtgtttaagtctg
ggcct
caaaataaacctatttttcaagcccaagaatctttaaaagattcaagtacgggccctgatttaaattcggataat
tttga
agaaggccgattacctttgcctggaggagtaagagaagatggtcaagaagttattgaagaaaatatttatacaga
tgaga
ttattaaaaacaaatcaatagaattttataagtacgcaatagaaaaaattgctgaattaaaatttgtatttggag
aagcc

```
ttagagaactacagaatatttaataaatcttcctacctaacatatctgtatgctttttgattttatttgctttt
ggcct
aggtgttggatttgtaagaaataatctcaaaaaacccgtgcaggaaaaagaaataattgataactcgttatcgat
aaatg
aaaataagaatgtcttttatgaaggtttaaatcaagatgataaaaagaaagttctcgataactcaaaaattattc
tctca
gataatgcagaaaaagttatttttcaggtgaagaaataaaaactgcttctccctccttagaaaaaatagaaaat
ttaat
taatacatggcttgttaacaaaagtaaatttctagcaggaaaaggtgaaattaatttatcaaagatagttcaaga
tgatt
tgattgatagattaaagaaggaaagagaacttgatattcaaaaaggtatctacaaaaatatcaatgctaatatcg
aaaat
attgtacttttaactcaaacggcatcaagaatatcagtatcagttgacttaaagtattcagaaaaaatattaaaa
ataga
tggggaattgataaatgaaacaactttcactccttttttgaaagttaaatatattttaggtttctcaaataactc
ctgga
aattagttgactacattagtggtgtttag
```

```
PROTEIN
LELPLDHFRLIGVSPSATSEEILRAFQLRLDKTPDEGFTYEVLTQRSELLRLTADLLTDPDSRRDYENLLLNGAS
GLDLS
SNREVAGLILLWESGSSKEAFKITRKALQPPQTPALGSSREADLTLLAALTSRDAAIQEQDQRSYSNAADFLQEG
IQLLQ
RMGKLGELRKTLEEDLVSLLPYRILDLLSRDLNDYDSHKKGLSMLENLIIKRGGLEGKNKSEYNDFLNQQEFESF
FQQIK
PFLTVQDQIDLFLELQKRGSSEAGFLAFLSLTAIGFARRKPAKLFEARKILKKLNLSGLDSMPLIGCLDLLLADV
EQSSA
RFLSSSDEKLRDWLNNYPGEKLEAICIFCKNWLENDVLVGYRDIDLKEIDLDSWFEDREIQEFIEQIEKKSNRTV
FKSGP
QNKPIFQAQESLKDSSTGPDLNSDNFEEGRLPLPGGVREDGQEVIEENIYTDEIIKNKSIEFYKYAIEKIAELKF
VFGEA
LENYRIFNKSSYLTYLYAFLILFAFGLGVGFVRNNLKKPVQEKEIIDNSLSINENKNVFYEGLNQDDKKKVLDNS
KIILS
DNAEKVIFSGEEIKTASPSLEKIENLINTWLVNKSKFLAGKGEINLSKIVQDDLIDRLKKERELDIQKGIYKNIN
ANIEN
IVLLTQTASRISVSVDLKYSEKILKIDGELINETTFTPFLKVKYILGFSNNSWKLVDYISGV*
```

DRAFT Prochlorococcus marinus sp. MIT9313 analysis files
Produced for the Joint Genome Institute Microbial Sequencing program.
N.B.: These pages subject to frequent change - work in progress.

http://genome.ornl.gov/cgi-bin/JGI_microbial/gene_viewer.cgi?org=pmar_mit&chr=18oct01&contig=Contig475&gene=2677

Version 18oct01 - Contig475 Gene 2677
Gene Finders
Strand = f
Stop Location = 398272
Stop Codon = taa

Gene Modeler	Start Location	Start Codon
Generation	396287	gtg
Glimmer	396287	gtg
Critica	396287	gtg

MRNA
gtggacctgccaatagatcatttccgcttgctgggtgtcagtccttcggcagacagtgaggcgattttgcgggcc
ttgga
gttgaggttggatcgctgccctgaccaaggtttcacccatgaggtcttaattcagcgggcagaattgttgcggct
ttcag
cagatttgctgactgatccgccacggcgtcaggcctatgagactgccttgttggagctcagtcgtgatcatccag
gtgag
accgccggtcttgatgtgtcacctagtagagaggtggcagggctgatcttgctgtttgaagcgaattcttctcat
gaggt
ttttcatctcgcctctcagggattgcaaccgccccagtccccgacgctaggtagcgaacgagaagctgacctcgc
tttgt
tgttggcactggcctgtcgggctgcagccgctgaggaacaggaacaacggcgttatgaagcagcagcgtctcttc
tgcat
gacgggatccagttgctgcagcggatgggcaagctctccgaagagtgccacaagcttgagaacgatttagatgcc
cttct
gccctatcgcattctcgacttattgagtcgggatcttggtgatcaggtttctcaccaggaaggactgcgcctact
tgaca
actttgtgagccagagaggaggtcttgagggaacggccccatcgcctgcacctggtggtcttgatcagtccgaat
ttgac
aacttcttcaagcagatcagaaagtttttaactgttcaggaacaggttgatcttttcctgcgctggcagcaagcc
ggatc
agcagatgcgggtttcctgggtgggttggctcttgctgctgttggattttcgcgtcggaagcctgaacgggtgca
ggaag
ctcggcagcacttagagaggcttcaactggatggatgcgacccgttgccgatgctgggttgcttggacctcttgc
tcgga
gatgtgggccgcgctcaggagcgttttctgcgcagtacagatcctcgagtgaaggactgtcttaacagccaccct
ggcga
tgaattggctgctttttgtgagtactgccgctcttggctgcgaggggacgtgcttcccggttatagggatgtgga
tgctg
aggccgttgatctagaggcttggtttgctgatcgggatgttcaggcttatgtggagcgcctggaacgcagcgaaa
atcgt
gcttcttctttaggtaaggccttctcaggatcgtctgtgaagcaacccttcccttgggcgcctcttgatcccgat
gggat
tttgcccctctctcttggtgggcctgatgttggtcaacctgcagctgatcagagctctgatgagtttgccagcga
tggta

```
tggcatggattgatcgtttagcagatctgccacgcccgacgcggccggtgctgatcggttcggttgtctttgcgg
ccctg
attgcagcctttgcaggcttcagtttgtttggccaacgtcctcgtacgtcagttagtacggctgctgatcagcct
caagt
cacagcacctcctacagccacactgcaagaggaggtcctcatgcctcaagtccctgtcagcgctgtggttgagcc
gctta
ctttggagcagccgaatgaggcacagctcaaaggcctgcttcaggcctggctcagcaacaaggcagtcgtgcttg
ccggt
ggcaagagtgatgcactgcctgaggtcgcaagagatccattggtgcagcgcgtggcgcaagagcgtgccagggat
gctgc
tttagctcagacccagaaggttgtggccagcatcagctctgtagaggtggtgagtcgaacgccgcagcgtattga
gctga
atgccgttgtgacctatcgcgatcaacgcgttgatgctgccggcaaggttgttgaccaaacgccccaaaaagatc
tctcg
gtgacttacatccttggtcgtgatcccgatcgttggcgcctgcatgaatacatcagcggcaaataa
```

```
PROTEIN
VDLPIDHFRLLGVSPSADSEAILRALELRLDRCPDQGFTHEVLIQRAELLRLSADLLTDPPRRQAYETALLELSR
DHPGE
TAGLDVSPSREVAGLILLFEANSSHEVFHLASQGLQPPQSPTLGSEREADLALLLALACRAAAAEEQEQRRYEAA
ASLLH
DGIQLLQRMGKLSEECHKLENDLDALLPYRILDLLSRDLGDQVSHQEGLRLLDNFVSQRGGLEGTAPSPAPGGLD
QSEFD
NFFKQIRKFLTVQEQVDLFLRWQQAGSADAGFLGGLALAAVGFSRRKPERVQEARQHLERLQLDGCDPLPMLGCL
DLLLG
DVGRAQERFLRSTDPRVKDCLNSHPGDELAAFCEYCRSWLRGDVLPGYRDVDAEAVDLEAWFADRDVQAYVERLE
RSENR
ASSLGKAFSGSSVKQPFPWAPLDPDGILPLSLGGPDVGQPAADQSSDEFASDGMAWIDRLADLPRPTRPVLIGSV
VFAAL
IAAFAGFSLFGQRPRTSVSTAADQPQVTAPPTATLQEEVLMPQVPVSAVVEPLTLEQPNEAQLKGLLQAWLSNKA
VVLAG
GKSDALPEVARDPLVQRVAQERARDAALAQTQKVVASISSVEVVSRTPQRIELNAVVTYRDQRVDAAGKVVDQTP
QKDLS
VTYILGRDPDRWRLHEYISGK*
```

172

Synechococcus sp. PCC7002
>gnl|jmarq_32049|Contig051302-306 Synechococcus sp. PCC 7002 unfinished fragment of genome Length = 107169

DNA:
>Synechococcus sp. PCC7002 Contig051302-306 position 55303..57453 reverse complement
GTGCGCATTCCGCTCGACTATTACCGCATCCTATGCGTCCCCGCCAAGGCAACCACTGCCCAAATTACCCAAGCC
TATCGCGATCGCCTCTCCCAATTTCCCCGTCGCGAACATAATGCCTTGGCCATTGAGGCCCGCAACCGGATTATC
GAGCAAGCCTTTGAGGTGTTATCCCAAACAGAAACCCGCGCCGTCTACGACCATGAGCTGTCGGGCAATATGTTT
CGTTCCCTCGTCCCCAGCCGTCCGAAACTGCCTTTTCCCGATCGCCCCTCCAGTGACACAGAGTTAGAAGCCCTG
ACAGCCCACCAACCAACCATTGACATCGCGGAAAAAGATTTACTGGGGGGACTGCTGTTACTCCTCGACCTGGGG
GAGTACGAATTAGTGCTGAAGTGGGCTGCCCCCTACCTCAAGGGCAAAGGCAAGCTGGTCAAGGAAGGGAAATTT
GGGGCCGTCGAAATCGTCGAGCAAGAACTACGGCTTTGTTTGGCCCTGGCCCACTGGGAATTGAGCCGGGAACAG
TGGCTCCAACAACATTATGAACAGGCGGCTCTCTCCGGTCAGAAGAGTCAAGAGCTATTGGTAGATGTGGCACAA
TTTGCAGACCTCCAACAGGAAATTCAAGGGGATCTCAATCGCCTCAGACCCTATCAAGTTCTAGAACTTCTGGCC
CTACCCGAATCAGAAACCCAAGAGCGACAACGGGGCTTACAACTGCTCCAGGAAATGTTGAGTGCTCGCGTGGGG
ATTGATGGCCAGGGGGACGATCAGTCGGGTCTAAGTATTGATGATTTTTTGCGCTTTATCCAGCAGTTACGCAGT
TATCTAACGGTGCAAGAACAGTTGGATCTCTTTGTGGCAGAATCAAAGCGACCTTCGGCGGCAGCGGCCTACCTA
GCGGTGTATGCTCTCTTGGCTGCTGGGTTTTCGCAACGGAAACCTGACCTGGTCGTGCAAGCCCAGACCCTATTA
AAACGCCTCGGCAAACGCCAGGATGTTTTCTTGGAGCAATCAATCTGCGCCTTACTTTTAGGTCAGCCGTCGGAA
GCCAATCAACTGTTAGAACAAAGTCAGGAACAGGAGGCGATCGCCTACATTCAAGAGCAGTCTGAGGGGGCACCG
GATCTACTCCCAGGCCTATGTCTCTACGGGGAACAGTGGCTGAAGACAGAGGTTTTTTCCCATTTCCGCGATCTC
CGGCAACGGCTTGAAGATGGCTCTGTTTCGTTGACGGCTTACTTCGCCGATCCTGAAGTGCAGCAATATCTTGAC
GATCTCCTCACGGAGGCTGTCCCCACACCCACACCACATCCAGACACAGAAAGTACAGCGGCCCCGTCGGAAAAG
CCACCGGAAACATTACAGTCAGAAACCGGTGTTTCGCCGCATCCCAGTCGTCCCGCCAAGGTTGATTCCTTTGAG
GATCTCGTCACTCAAACTCCCGCTACAGTTCCCCCGGCACCGCCTTCTCCTGGTGTAGCACCTGTAACTGCGGCA
TTAAACCCAGACCCGGAAGCGTCTTCTGCTTCGTCAAAATCAGTTTCGTCAAAAAAGTCTATCGGGCCTTGGGGG
GCGATCGCCGCTATCGTGGGGAGTGTTTTGCTGGTCGTGGGCCTGGTGCGAATTTTGTCTGGCCTAACTACCCAG
GAACCCTTACAGGTCACCCTCAACGGTGAGCCACCCCTAACGATCCCCAGCTTAGACACCGCCGAGGCAAATAAT
AATCCGGAGAATGGAGCGACCGATACAACGACAACGCCTGCGCTCAATGAGGCGATCGCCGCTGAGGTGATTCAA
ACTTGGTTTGAGAGTAAAGCTAGAGCCTTTGGCCAAGACCGTGATTTGGCGGCTCTAGAAAATATTTTGGCAGAA
CCGTCCCTGTCCCGCTGGCGCAGTAGTGCCCAGGCCGTCCGCAGCGCTGGTACCTACCGCACCTATGACCACAGT
TTGACCATTGAAACGGTGAGCTTCAACCCAGACCAACCCAATGTGGCGACCGTTGAGGCCCAGGTGCAGGAAAAG
GCAGATTATTACCGGGCGAATGGGGAACGCGATCCCGGCCAGTCCTATGATTCTGACCTGCGTGTCCGCTACAGC
TTGGTGCGCCAAGGCGATCGCTGGTTGATTCGTTCTTCCCAAACCCTGTAA

Protein:
>Scc_7002_Sequence 1 ORF:57453.. 55303 Frame -2
MRIPLDYYRILCVPAKATTAQITQAYRDRLSQFPRREHNALAIEARNRIIEQAFEVLSQTETRAVYDHELSGNMF
RSLVPSRPKLPFPDRPSSDTELEALTAHQPTIDIAEKDLLGGLLLLLDLGEYELVLKWAAPYLKGKGKLVKEGKF
GAVEIVEQELRLCLALAHWELSREQWLQQHYEQAALSGQKSQELLVDVAQFADLQQEIQGDLNRLRPYQVLELLA
LPESETQERQRGLQLLQEMLSARVGIDGQGDDQSGLSIDDFLRFIQQLRSYLTVQEQLDLFVAESKRPSAAAAYL
AVYALLAAGFSQRKPDLVVQAQTLLKRLGKRQDVFLEQSICALLLGQPSEANQLLEQSQEQEAIAYIQEQSEGAP
DLLPGLCLYGEQWLKTEVFSHFRDLRQRLEDGSVSLTAYFADPEVQQYLDDLLTEAVPTPTPHPDTESTAAPSEK
PPETLQSETGVSPHPSRPAKVDSFEDLVTQTPATVPPAPPSPGVAPVTAALNPDPEASSASSKSVSSKKSIGPWG
AIAAIVGSVLLVVGLVRILSGLTTQEPLQVTLNGEPPLTIPSLDTAEANNNPENGATDTTTTPALNEAIAAEVIQ
TWFESKARAFGQDRDLAALENILAEPSLSRWRSSAQAVRSAGTYRTYDHSLTIETVSFNPDQPNVATVEAQVQEK
ADYYRANGERDPGQSYDSDLRVRYSLVRQGDRWLIRSSQTL

```
LOCUS       AF421196                2469 bp    DNA     linear   BCT 18-OCT-
2001
DEFINITION  Synechococcus sp. PCC 7942 cell division protein Ftn2 gene,
            complete cds.
ACCESSION   AF421196
VERSION     AF421196.1  GI:16226083
KEYWORDS    .
SOURCE      Synechococcus sp. PCC 7942.
  ORGANISM  Synechococcus sp. PCC 7942
            Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
REFERENCE   1  (bases 1 to 2469)
  AUTHORS   Koksharova,O.A. and Wolk,C.P.
  TITLE     Two novel genes, one bearing a DnaJ motif, are involved in
control
            of cyanobacterial cell division
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2469)
  AUTHORS   Koksharova,O.A. and Wolk,C.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Plant Research Laboratory, Michigan
State
            University, DOE Plant Research Laboratory, East Lansing, MI
48824,
            USA.
FEATURES             Location/Qualifiers
     source          1..2469
                     /organism="Synechococcus sp. PCC 7942"
                     /strain="PCC 7942"
                     /db_xref="taxon:1140"
     CDS             319..2214
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein Ftn2"
                     /protein_id="AAL16071.1"
                     /db_xref="GI:16226084"

/translation="MRIPLDYYRILCVGVQASADKLAESYRDRLNQSPSHEFSELALQ

ARRQLLEAAIAELSDPEQRDRYDRRFFQGGLEAIEPSLELEDWQRIGALLILLELGEY

DRVSQLAEELLPDYDASAEVRDQFARGDIALAIALSQQSLGRECRQQGLYEQAAQHFG

RSQSALADHQRFPELSRTLHQEQGQLRPYRILERLAQPLTADSDRQQGLLLLQAMLDD

RQGIEGPGDDGSGLTLDNFLMFLQQIRGYLTLAEQQLLFESEARRPSPAASFFACYTL

IARGFCDHQPSLIHRASLLLHELKSRMDVHIEQAIASLLLGQPEEAEALLVQSQDEET

LSQIRALAQGEALIVGLCRFTETWLATKVFPDFRDLKERTAPLQPYFDDPDVQTYLDA

IVELPSDLMPTPLPVEPLEVRSSLLAKELPTPATPGVAPPPRRRRRDRSERPARTAKR

LPLPWIGLGVVVVLGGGTGVWAWRSRSNSTPPTPPPVVQTLPEAVPAPSPAPVTVALD

RAQAETVLQNWLAAKAAALGPQYDRDRLATVLTGEVLQTWQGFSSQQANTQLTSQFDH
```

74

KLTVDSVQLSDGDQRAVVQAKVDEVEQVYRGDQLLETRRDLGLVIRYQLVRENNIWKI
ASISLVR"

BASE COUNT 493 a 712 c 712 g 552 t

ORIGIN

```
        1 cttgccgact aaaggctaag catcgccatt ccttagatta aagcagtctg tcggcggcgc
       61 tgtgccggtt aacaccagtc tgtcgctgac agcggtgcct ttctggggct tgcctgtggg
      121 gcgagtaacc gatcgctggg ataagagttg gtgcttctgg ctctcaagaa tagggttttc
      181 cgtcgcgtat tcccgatcac atccccctgt gtctgctacg gagataacgc cgatcactca
      241 acagaattgg taagttgacg gtcaagttgg gatgatgaag tcggctcaag ctggcgatcc
      301 ggatctggtg ggtgttctgt gcgtattcct ctcgattact accgaattct ctgtgttggc
      361 gtgcaagcct cggcagacaa acttgccgaa agctaccgcg atcgcctcaa ccaatcgccc
      421 tcccatgagt tttcagagct ggcattgcag gcgcggcggc aactcctcga agcagcgatt
      481 gctgagctga gtgatcccga acagcgcgat cgctacgatc gccgcttttt tcagggcggt
      541 ctggaagcga ttgaaccaag cctagaactc gaagactggc agcgaattgg agccctgctg
      601 atcctgctgg aattggggga atacgatcgc gtttcgcaac tggctgagga actcctgcca
      661 gactacgacg cgagcgcaga agtacgcgat cagttcgcgc ggggtgatat cgccttggcg
      721 atcgcactat cccagcaatc cctcggtcga gaatgccgtc agcagggtct gtacgaacag
      781 gccgcccagc actttggccg cagccagtct gccctagccg atcatcagcg ctttcctgaa
      841 ctgagtcgaa ccctgcacca agaacaagga cagctacggc cctatcgcat tttggagcgg
      901 ttggcccagc ccttgactgc cgatagcgat cgccagcagg gtttgctgtt gttgcaggcg
      961 atgttggacg accggcaggg cattgaaggc cctggggatg atggctcggg gctgaccctt
     1021 gataactttt tgatgtttct ccagcaaatt cgcggctatc tgaccctggc tgaacagcag
     1081 ttgctgtttg aatcggaagc gcgtcggccc tcgccggctg cgagcttttt tgcctgctac
     1141 accctgattg cgcggggctt ttgcgatcac caaccctcgt tgatccatcg cgccagcttg
     1201 ctcttgcatg aactcaagag ccgcatggat gtgcacatcg aacaggcgat cgccagccta
     1261 ttgctcggac agcccgaaga agctgaggcg ctactcgtcc agagccaaga tgaggaaacc
     1321 ctcagccaaa tccgtgccct agcccaaggg gaagccctga tcgtcggttt gtgccgattc
     1381 acggaaacct ggctagcgac caaggtattt ccggatttcc gcgacctcaa ggaaaggact
     1441 gcgccgctgc agccctactt tgacgacccc gatgtccaga cctatctgga tgcgatcgtg
     1501 gagttgccgt ccgatttgat gccaacgccg ctacccgttg agccgcttga ggtgcgatcg
     1561 tcgttgctgg ccaaggaact gccgacccca gcaacgcctg gtgtagctcc accccctcgc
     1621 cgccgtcgcc gcgatcgctc cgaacgtcct gctcgcacgg ccaaacgctt gcccttgccc
     1681 tggattggtt tgggggttgt ggtggttctc ggcggtggaa caggggtttg ggcttggcga
     1741 tcgcgttcca attccacccc gccgaccccg ccccccgtgg ttcaaacgct gcctgaggcg
     1801 gtacctgccc cttcgcccgc gccagttacc gttgccctcg atcgggctca ggctgaaact
     1861 gtgttgcaaa actggttggc cgctaaagct gcagccttgg ggcctcaata cgatcgcgat
     1921 cgcttagcga cggtgctgac cggtgaggtt ctgcagactt ggcagggttt ttctagccag
     1981 caggccaaca cccagctcac atcacagttc gatcacaagt taaccgtcga ctcagttcag
     2041 ctcagtgacg gtgatcaacg agcagtagtc caagccaagg tcgatgaagt tgagcaggtc
     2101 tatcgaggcg accagctgct cgaaacgcgc cgagatttgg gcttggtgat ccgctaccag
     2161 ctcgtgcgcg agaacaacat ctggaaaatt gcttcgatta gtttggtgcg ctaggaattc
     2221 gcaaggggtg aacccccctgc ggtcttttct gtagatcccc tagagcgatc gcagaatgtt
     2281 cagcgattcc tggatgtgcg cttgggcatt caagagtgaa tcaaaaatgt ggcgcacctt
     2341 gccctctttg tcgatcacat aagtgacgcg acccggaatc acaaacaggg ttttgggcac
     2401 gccataggtt tgacggaggc gatcgcctgc atcgctcagc agttggaagg gcaagttgta
     2461 tttctgggc
```

//

```
LOCUS       AF421196_1              631 aa            linear   BCT 18-OCT-
2001
DEFINITION  cell division protein Ftn2 [Synechococcus sp. PCC 7942].
ACCESSION   AAL16071
VERSION     AAL16071.1  GI:16226084
DBSOURCE    locus AF421196 accession AF421196.1
KEYWORDS    .
SOURCE      Synechococcus sp. PCC 7942.
  ORGANISM  Synechococcus sp. PCC 7942
            Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
REFERENCE   1  (residues 1 to 631)
  AUTHORS   Koksharova,O.A. and Wolk,C.P.
  TITLE     Two novel genes, one bearing a DnaJ motif, are involved in
control
            of cyanobacterial cell division
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 631)
  AUTHORS   Koksharova,O.A. and Wolk,C.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Plant Research Laboratory, Michigan
State
            University, DOE Plant Research Laboratory, East Lansing, MI
48824,
            USA.
COMMENT     Method: conceptual translation supplied by author.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Synechococcus sp. PCC 7942"
                     /strain="PCC 7942"
                     /db_xref="taxon:1140"
     Protein         1..631
                     /product="cell division protein Ftn2"
     CDS             1..631
                     /coded_by="AF421196.1:319..2214"
                     /transl_table=11
ORIGIN
        1 mripldyyri lcvgvqasad klaesyrdrl nqspshefse lalqarrqll eaaiaelsdp
       61 eqrdrydrrf fqggleaiep sleledwqri gallillelg eydrvsqlae ellpdydasa
      121 evrdqfargd ialaialsqq slgrecrqqg lyeqaaqhfg rsqsaladhq rfpelsrtlh
      181 qeqgqlrpyr ilerlaqplt adsdrqqgll llqamlddrq giegpgddgs gltldnflmf
      241 lqqirgyltl aeqqllfese arrpspaasf facytliarg fcdhqpslih raslllhelk
      301 srmdvhieqa iaslllgqpe eaeallvqsq deetlsqira laqgealivg lcrftetwla
      361 tkvfpdfrdl kertaplqpy fddpdvqtyl daivelpsdl mptplpvepl evrssllake
      421 lptpatpgva pppprrrrdr serpartakr lplpwiglgv vvvlgggtgv wawrsrsnst
      481 pptpppvvqt lpeavpapsp apvtvaldra qaetvlqnwl aakaaalgpq ydrdrlatvl
      541 tgevlqtwqg fssqqantql tsqfdhkltv dsvqlsdgdq ravvqakvde veqvyrgdql
      601 letrrdlglv iryqlvrenn iwkiasislv r
```

>gi|17131676|dbj|AP003590.1|AP003590 Nostoc sp. PCC 7120 DNA, complete genome, section 10/19 Length = 333500

nt 213526 .. 211130
Frame = -2
DNA:
>AP003590 213526 .. 211130 reverse complement

```
ATTATGTTGATCACGGTGCAGGGGAAGTACGCTGTGCGAATTCCGCTAGATTACTACCGAATTTTAGGGCTACCG
TTAGCGGCAAGTGATGAACAACTGCGACAAGCATACAGCGATCGCATTGTCCAATTGCCGCGACGGGAGTATTCT
CAAGCAGCAATTGCTTCCCGTAAACAACTTATAGAAGAAGCTTACGTGGTTTTATCAGATCCAAAGGAACGCAGC
AGTTATGACCAGCTGTATCTTGCTCACGCCTACGACCCAGACAACGCGGCTACAACCAAAGTGGCAGTGGAAAAT
CGTGGGGACAGCAACAATGGTCATTTCGATGTCCAAAGCCTGAGCATCGAAGTTTCCTCCGAGGAATTAATTGGT
GCTTTATTAATTTTGCAAGAGTTGGGAGAGTATGAACTCGTACTCAAGTTAGGTCGTAATTACTTAGGTAATCAA
AACGGCACAGCATCCACCAGAAATGGCAATCATCGCACGCCTGAAGAATTTCTCGATAGTTCTGAACGTCCAGAT
ATTCTCTTGACTGTTGCTTTGGCCTCATTAGAATTAGGGCGGGAACAATGGCAACAAGGCCACTATGAAAACGCT
GCTTTGTCTTTAGAGACTGGGCAAGAAGTGCTGTTTAGTGAAGGCATCTTCCCCAGCGTCCAGGCAGAAATTCAG
GCTGATCTTTACAAATTACGCCCTTATAGAATTTTAGAATTACTTGCCTTACCCCAGGAAAAAACCATTGAACGC
CACCAAGGGCTGGATCTATTACAAAGCATCTTAGACGATCGCGGTGGCATTGATGGTACAGGCAATGATCAATCA
GGCTTAAACATTGATGACTTCCTCCGATTCATCCAGCAATTACGCCACCACTTAACAGTGGCTGAACAACATAAG
TTGTTTGATGGTGAAAGCAAACGCCCTTCGGCTGTGGCTACATACTTAGCTGTTTATGCTTCCATCGCCAGAGGA
TTCACCCAACGCCAGCCCGCTTTAATTCGTCATGCCAAGCAAATTCTGATGCGTTTGTCTAAGCGGCAAGATGTG
CATTTAGAGCAGTCCCTGTGTGCGCTATTACTAGGGCAAACTGAAGAAGCCACGCGAGTTTTAGAACTGAGCCAA
GAATACGAAGCTTTAGCCTTAATTCGAGAAAAATCTCAAGATTCACCCGATTTACTGCCAGGTTTGTGCTTATAT
GCCGAACAATGGCTGCAAAATGAAGTTTTCCCCCATTTCCGCGATTTGTCCAGACAGCAAGCTTCCCTGAAAGAT
TACTTTGCTAATCAACAAGTACAAGCGTATTTAGAAGCCTTGCCCAACGACGCGGAAACCACTAATGAATGGGCT
GTAATTAACCGCCAATCGTTTTCTCAACCCAGGGGCAATTCTTACTCTGGAGGAACGCCAGTCGCCAAACGTCCC
GTAGGGAAGGCGAACAGGCCAGGAGAAGCGTCCACAAGACCAGTTCCCCAACGTAGTCATCCATCAGAAGTAAAT
CGGCAGTTTCATCAAAACAGAACCCCTGATCCCGAATTACCAGAAACATCAAACCACAGAAGACCAGAGTCTTCA
AATTTTACAACTGCTAGAGAAAATATATCGACCACAGATGCTTACACTGACAATTATCCACCAGAGATCCCTGTA
GAACGCGCCAGCAGACCTGTTCAGCCGGGGGTAAGTGGTTATACCCAATCGACCCCTCCACGGCAAACTCCTAAA
CGCAGGAGACGCAAGAAGCCACAGGCAGTTGTCAACAGAGGACACAGTATTCATCAGCAACGCCAACCCTCACCT
AGCACTCTAGGCCGGAAAACAAGATTACTTTGGATAGTTTTGGGTTCTTTGGGTGGGATATTATTGTTCTGGCTG
ATAGTCTCAACGACTTTTGGGTGGTTAAAGAATGTATTCTTCCCAGCACCATCTTTACAAGGTGAGCAATTATCG
ATTCAGATTAGTCAACCACCTTTAGAGATTCCTGACAAAAATGCCCAGATACAATCCCCAGAGGTGAGTCTCACA
GAAGAAACGGCAAGGAAAATAATTGAAAATTGGTTGGCTACCAAAGCTAGTGCTTTAGGCGCTGAACATAAAATT
GAGAGTTTAAACGAGATTTTAACTGGTTCAGCGTTATCTCAATGGCGGCTAATTGCCTTGCAAGATAAAGCAGAC
AATCGTCATCGAGAATACAGTCATAGTGTCAAGGTAGACTCCATCAGTAAATCTGACATAGATCCCAATCGTGCA
AGTGTGGGGGCTACAGTCAGAGAGTTAACCCAATTTTATGAGAATGGGCAAAAAGGGAAGTCTTCTGACGAAAGA
TTACGTGTACGCTATGAATTGATTCGACAAGATGATATTTGGCGGATTCAGAGGATGTCAGCCGCTATAAATTAA
```

Protein:

```
LOCUS       BAB74406                798 aa            linear   BCT 28-NOV-
2001
DEFINITION  ORF_ID:all2707~hypothetical protein [Nostoc sp. PCC 7120].
ACCESSION   BAB74406
VERSION     BAB74406.1  GI:17131800
DBSOURCE    locus AP003590 accession AP003590.1
KEYWORDS    .
SOURCE      Nostoc sp. PCC 7120.
  ORGANISM  Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
```

```
            Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
            Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
            Yasuda,M. and Tabata,S.
  TITLE     Complete genomic sequence of the filamentous nitrogen-fixing
            cyanobacterium Anabaena sp. strain PCC 7120
  JOURNAL   DNA Res. 8 (5), 205-213 (2001)
  MEDLINE   21595285
   PUBMED   11759840
REFERENCE   2  (residues 1 to 798)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyanobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..798
                     /organism="Nostoc sp. PCC 7120"
                     /db_xref="taxon:103690"
                     /note="synonym:Anabaena sp. PCC7120"
     Protein         1..798
                     /name="ORF_ID:all2707
                     hypothetical protein"
     CDS             1..798
                     /gene="all2707"
                     /coded_by="complement(AP003590.1:211130..213526)"
                     /transl_table=11
ORIGIN
        1 mlitvqgkya vripldyyri lglplaasde qlrqaysdri vqlprreysq aaiasrkqli
       61 eeayvvlsdp kerssydqly lahaydpdna attkvavenr gdsnnghfdv qslsievsse
      121 eligallilq elgeyelvlk lgrnylgnqn gtastrngnh rtpeefldss erpdilltva
      181 laslelgreq wqqghyenaa lsletgqevl fsegifpsvq aeiqadlykl rpyrilella
      241 lpqektierh qgldllqsil ddrggidgtg ndqsglnidd flrfiqqlrh hltvaeqhkl
      301 fdgeskrpsa vatylavyas iargftqrqp alirhakqil mrlskrqdvh leqslcalll
      361 gqteeatrvl elsqeyeala lireksqdsp dllpglclya eqwlqnevfp hfrdlsrqqa
      421 slkdyfanqq vqaylealpn daettnewav inrqsfsqpr gnsysggtpv akrpvgkanr
      481 pgeastrpvp qrshpsevnr qfhqnrtpdp elpetsnhrr pessnfttar enisttdayt
      541 dnyppeipve rasrpvqpgv sgytqstppr qtpkrrrrkk pqavvnrghs ihqqrqpsps
      601 tlgrktrllw ivlgslggil lfwlivsttf gwlknvffpa pslqgeqlsi qisqppleip
      661 dknaqiqspe vslteetark iienwlatka salgaehkie slneiltgsa lsqwrlialq
      721 dkadnrhrey shsvkvdsis ksdidpnras vgatvreltq fyengqkgks sderlrvrye
      781 lirqddiwri qrmsaain
//
```

```
LOCUS       NP_486747                798 aa            linear   BCT 28-NOV-
2001
DEFINITION  hypothetical protein [Nostoc sp. PCC 7120].
ACCESSION   NP_486747
VERSION     NP_486747.1  GI:17230199
DBSOURCE    REFSEQ: accession NC_003272.1
KEYWORDS    .
SOURCE      Nostoc sp. PCC 7120.
  ORGANISM  Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
            Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
            Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
            Yasuda,M. and Tabata,S.
  TITLE     Complete genomic sequence of the filamentous nitrogen-fixing
            cyanobacterium Anabaena sp. strain PCC 7120
  JOURNAL   DNA Res. 8 (5), 205-213 (2001)
  MEDLINE   21595285
   PUBMED   11759840
REFERENCE   2  (residues 1 to 798)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyanobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to
final
            NCBI review. The reference sequence was derived from BAB74406.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..798
                     /organism="Nostoc sp. PCC 7120"
                     /db_xref="taxon:103690"
     Protein         1..798
                     /name="hypothetical protein"
     CDS             1..798
                     /gene="all2707"
                     /coded_by="complement(NC_003272.1:3300430..3302826)"
                     /transl_table=11
ORIGIN
        1 mlitvqgkya vripldyyri lglplaasde qlrqaysdri vqlprreysq aaiasrkqli
       61 eeayvvlsdp kerssydqly lahaydpdna attkvavenr gdsnnghfdv qslsievsse
      121 eligallilq elgeyelvlk lgrnylgnqn gtastrngnh rtpeefldss erpdilltva
      181 laslelgreq wqqghyenaa lsletgqevl fsegifpsvq aeiqadlykl rpyrilella
      241 lpqektierh qgldllqsil ddrggidgtg ndqsglnidd flrfiqqlrh hltvaeqhkl
      301 fdgeskrpsa vatylavyas iargftqrqp alirhakqil mrlskrqdvh leqslcalll
      361 gqteeatrvl elsqeyeala lireksqdsp dllpglclya eqwlqnevfp hfrdlsrqqa
      421 slkdyfanqq vqaylealpn daettnewav inrqsfsqpr gnsysggtpv akrpvgkanr
      481 pgeastrpvp qrshpsevnr qfhqnrtpdp elpetsnhrr pessnfttar enisttdayt
      541 dnyppeipve rasrpvqpgv sgytqstppr qtpkrrrrkk pqavvnrghs ihqqrqpsps
```

79

```
601 tlgrktrllw ivlgslggil lfwlivsttf gwlknvffpa pslqgeqlsi qisqppleip
661 dknaqiqspe vslteetark iienwlatka salgaehkie slneiltgsa lsqwrlialq
721 dkadnrhrey shsvkvdsis ksdidpnras vgatvreltq fyengqkgks sderlrvrye
781 lirqddiwri qrmsaain
```

DRAFT Nostoc punctiforme analysis files
Produced for the Joint Genome Institute Microbial Sequencing program.
N.B.: These pages subject to frequent change - work in progress.
http://genome.ornl.gov/cgi-bin/JGI_microbial/gene_viewer.cgi?org=npun&chr=31may01&contig=Contig493&gene=84

Version 31may01 - Contig493 Gene 84
Gene Finders
Strand = r
Stop Location = 105061
Stop Codon = TAA

Gene Modeler	Start Location	Start Codon
Generation	107367	GTG
Glimmer	107367	GTG
Critica	107367	GTG

MRNA
GTGCGAATTCCGCTAGATTACTACCGAATTTTAGGACTACCGTTAGCGGCAAGTGAAGAACAATTGCGACAGGCA
TACAG
CGATCGCATTGTACAATTGCCACGACGTGAGTATTCTCAGGCAGCAATTTCTTCTCGTAAACAACTCATAGAAGA
AGCTT
ACGTGGTTTTATCAGATCCAAAACAACGCAGTACCTACGATCAGCTTTATCTTGCCCACGCCTATGACCCTGATA
ACCTT
GCTGCTGCCGCAGTAGCACAGGAAAATCGTACAGAAAGCACCAAAAGGGGTAGTGATACCCAGAGTCTTGGTATA
GAAAT
TACCCAAGACGAATTAGTTGGCGCTTTATTAATTTTGCAAGAGTTGGGTGAATACGAACTTGTATTGAAACTAGG
TCGTC
CGTACCTAGTAAATAAAAATAGTGCTACAAGTTCAAGAAAAAGCAATAACTTAGCAGATGAAGAAATTTATGAAA
GTGCT
GAACACCCAGATGTCGTTCTCACTGTTGCTCTTGCCTGTCTAGAATTAGGTCGGGAACAGTGGCAGCAAGGTCAC
TACGA
AAATGCCGCCATATCCCTAGAAACTGGTCAAGAGCTGCTAGTACGTGAAGGTTTGTTCTCCAGTATCCAGGCAGA
AATTC
AGGCTGATCTTTACAAATTGCGGCCATATCGAATTTTGGAGTTGCTCGCATTACCTCAAGAAAAGACTGCCGAAC
GAAGC
CAAGGCTTAGAATTATTGCAAAATCTCTTAGAAGATCGTGGCGGGATTGATGGCACGAACAATGATGAATCGGGT
TTAAA
CATAGATGACTTTCTGCGATTTATCCAGCAGTTACGCAACCACTTAACAGTTGCAGAACAGCACAAGTTATTTGA
AGCTC
AAAGCAAACGTTCTTCTGCTGTTGCCACTTACTTAGCTGTTTATGCCTTGATAGCGCGAGGATTTGCTCAACGGC
AACCT
GCTTTAATTCGTCAAGCAAGACAAATGCTCGTGCGTCTGGGCAAGCGCCAAGATGTACATTTAGAACAGTCGCTA
TGTGC
CTTACTTTTGGGGCAAACTGAAGAAGCAACTCGTGTTTTAGAACTTAGTCAGGAGTACGAAGCTTTAGCTTTTAT
TCGGG
AAAAATCTCAGGACTCTCCAGATTTGTTACCGGGTCTGTGTTTATATGCAGAACAGTGGCTGCAACACGAAGTCT
TTCCC
CATTTTCGAGATTTAGCAAACCAGCAAGCTTTCCTAAAAGATTACTTTGCTAACCAACAGGTGCAAGCTTATTTA
GAAGC
ACTGCCAACTGATGCCCAAACAACTAATGAATGGGCTGTAATTAACCCCCAGTATTTTCCCCAGGCCAAGGCAAA
GAATA
CTCATTTTCATAACAATTCAACTAAAACTTCAGCGTCATTTAATCACAGCAGAGTACCTAACCCAGATTTGCCAG
AAACA

CCAACAAAAGAAACCTCTGAATATCCAAACTTCTCACCACCTATGTGGAGTTCATCTGGAAGTATAAAATCAGAG
GTTCC
TGCTGCTGAAAGGATGAGCAGAGGTACTAATCAGCATTTGAACGGTTCAGCTAAGAGTGCTGCATCTGGTCATAA
CCAAA
AGCGTAGGCGGAGAAAACCTACTCCATCTGCTAGCCGAGAGCGTATACCAGATAATCGTCCTCATTCTCGTCGTC
CCCGA
AGGCGGCGAACTTTTGCGAACACCATAGAAGGTAAAACACGGCTGGTATGGAGAGTGTTTATTTCTTTGGTGAGC
ATATT
AGTTTTTTGGGTATTAGCCACAACAACTTTTGGATGGTTAAAAAATCTGTTTTTTCCTCAACCTTCTCCGCCTGA
TCTAC
AGTTGTTTGTACAAATAAACCAACCACCGTTACCTATTCCCGATCCAAATAGAAAACCAGAATCAGAAGAAGGCC
CTTTA
ACAAATGCAGAGGCAGAAGAAGTTATTCACACTTGGTTATCTACCAAAGCCGCAGCTTTAGGGCCCAATCATGAG
ATTAA
TAATTTAGAGCAAATTTTAACTGGTTCAGCTTTATCTCAATGGCGACTGATTGCTCAACAGAATAAGTTAGACAA
TCGCT
ACCGCAAGTTCGACCATAGTTTGAAGATAGAATCTGTTGAGAAAATTGGTTTATTTGCAGATCGTGCCGCAGTAG
AAGCT
ACGGTCAAAGAAGTGACGCAGTTATATGAAAATAATCAGTTTAAAAACTCTTCTAACGATAAATTAAGAGTTCGG
TATGA
CTTGATTCGAGAACGAGGTAAATGGCGTATTCAGAGTACATCTGTTGTAAATCAATTCACCAGATAA

PROTEIN
VRIPLDYYRILGLPLAASEEQLRQAYSDRIVQLPRREYSQAAISSRKQLIEEAYVVLSDPKQRSTYDQLYLAHAY
DPDNL
AAAAVAQENRTESTKRGSDTQSLGIEITQDELVGALLILQELGEYELVLKLGRPYLVNKNSATSSRKSNNLADEE
IYESA
EHPDVVLTVALACLELGREQWQQGHYENAAISLETGQELLVREGLFSSIQAEIQADLYKLRPYRILELLALPQEK
TAERS
QGLELLQNLLEDRGGIDGTNNDESGLNIDDFLRFIQQLRNHLTVAEQHKLFEAQSKRSSAVATYLAVYALIARGF
AQRQP
ALIRQARQMLVRLGKRQDVHLEQSLCALLLGQTEEATRVLELSQEYEALAFIREKSQDSPDLLPGLCLYAEQWLQ
HEVFP
HFRDLANQQAFLKDYFANQQVQAYLEALPTDAQTTNEWAVINPQYFPQAKAKNTHFHNNSTKTSASFNHSRVPNP
DLPET
PTKETSEYPNFSPPMWSSSGSIKSEVPAAERMSRGTNQHLNGSAKSAASGHNQKRRRRKPTPSASRERIPDNRPH
SRRPR
RRRTFANTIEGKTRLVWRVFISLVSILVFWVLATTTFGWLKNLFFPQPSPPDLQLFVQINQPPLPIPDPNRKPES
EEGPL
TNAEAEEVIHTWLSTKAAALGPNHEINNLEQILTGSALSQWRLIAQQNKLDNRYRKFDHSLKIESVEKIGLFADR
AAVEA
TVKEVTQLYENNQFKNSSNDKLRVRYDLIRERGKWRIQSTSVVNQFTR*

>*Synechocystis* sp. strain PCC6803 D63999:2314780-2316924 complement
GTGTTTATCCCCCTCGACTTTTATCGTATTTTAGGCATTCCTCCCCAGAGTGGTGGGGAA
ACCATTGAGCAGGCCTACCAAGATCGCCTTTTACAATTACCCCGGCGAGAATTTAGTGAC
GCCGCAGTTACTCTCCGCAATCAATTACTGGCGATCGCCTATGAAACCCTGAGGGATCCG
GAAAAACGTCAGGCATACGACCAAGAATGGTGGGGAGCCATGGATGAAGCCCTGGGGGAG
GCCTTACCCCTCACTACCCCGGAGTTGGAATGTAGCCCAGAGCAAGAAATTGGAGCCCTG
TTGATCCTGTTGGATTTGGGGGAATACGAACTCGTGGTTAAGTATGGTGAGCCAGTACTC
CACGATCCCAACCCTCCGGCGGGAGGCCTGCCCCAGGACTATTTGCTTTCGGTAATTTTG
GCCCACTGGGAACTGAGCCGGGAACGTTGGCAACAACAGCAGTATGAATTTGCCGCCACC
GCCAGTCTTAAGGCCCTAGCTCGGTTGCAACAGGATAATGACTTCCCCGCCTTGGAAGCA
GAAATTCGTCAGGAACTATACCGTCTGCGACCCTACCGTATCCTCGAACTTTTGGCTAAG
GAGGGGCAAGGGGAGGAGCAACGTCAGCAGGGTCTAGCTCTGTTGCAAGCGATGGTGCAG
GACCGGGGCGGCATTGAAGGTAAGGGGGAAGATTATTCCGGATTGGGAAATGATGACTTT
CTAAAATTCATCCACCAACTACGCTGTCACCTCACAGTGGCCGAGCAAAACGCCCTATTT
TTGCCCGAAAGTCAACGGCCATCTTTAGTAGCAAGCTATTTGGCAGTACATAGTCTGATG
GCTGAGGGAGTGAAGGAACAGGACCCCATGGCCATTGTCGAAGCAAAATCTTTGATTATA
CAGTTGGAAAATTGTCAAGATTTGGCCCTAGAAAAGGTAATTTGTGAATTATTATTGGGT
CAAACGGAAGTTGTTCTGGCGGCGATCGACCAGGGAGATCCGAAAATAGTAGCTGGCCTC
GAATCTAAGTTAGCGACGGGGGAAGACCCCTTAACTGCTTTTTATACTTTCACTGAGCAG
TGGCTAGAGGAAGAAATTGTCCCCTACTTTAGGGATCTTTCTCCGGAGACCCTTTCCCCC
AAGGCCTATTTCAATAATCCCTCCGTTCAGCAGTATCTAGAACAACTAGAGCCGGATTCC
TTCACCACTGACAATTCTTTTGCCTCCCCTGCCCTCCTTAGCACCGCAACGGAATCGGAA
ACTCCCATGGTACATAGTTCCGCCGCCCTTCCCGATCGCCCTTTGACCTCCACCGTTCCC
TCACGACGGGGACGCAGTCCAAGACGTTCCCGAGACGATGTTTTCCCCAGCGCCGACAAT
TCCAGTGGTTTGGCCGTCACCACCCTATCTCCGGCGATCGCCTACGACACCCACTCCTTG
GGCACCAACGGTATTGGCGGGGATAGCACTAGCAACGGTTTTTCCAGTAACTCCGCCCCA
GAATCCACCAGTAAACATAAATCTCCCCGGCGACGCAAAAAACGGGTGACCATCAAGCCG
GTGCGCTTCGGCATTTTTCTGCTTTGCCTAGCAGGCATTGTGGGGGGGGCAACTGCCCTA
ATTATCAATCGTACTGGCGATCCCCTAGGTGGGTTGCTAGAAGACCCCCTAGATGTTTTC
CTGGACCAACCTTCAGAATTTATCCCCGATGAAGCCACGAGCCGGAATTTGATTCTCAGT
CAACCCAACTTCAATCAGCAAGTGGGTCAGATGGTAGTACAAGGCTGGCTTGATAGTAAA
AAGTTAGCCTTTGGCCAAAACTACGATGTCGGGGCATTGCAGAGTGTTTTAGCCCCCAAT
CTCCTTGCCCAACAACGGGGTCGGGCCCAACGGGATCAAGCCCAAAAGGTCTATCACCAA
TACGAACACAAGTTGCAGATTTTAGCCTATCAAGTTAACCCCCAAGACCCCAACCGAGCC
ACCGTTACTGCCCGGGTAGAAGAAATTAGCCAGCCCTTTACCCTAGGTAATCAACAGCAG
AAGGGCTCCGCCACCAAAGATGACTTGACTGTGCGCTATCAGCTAGTACGACACCAAGGG
GTTTGGAAAATTGACCAAATACAAGTGGTAAATGGCCCCCGTTAG

LOCUS NP_441990 714 aa linear BCT 23-OCT-2001
DEFINITION unknown protein [Synechocystis sp. PCC 6803].
ACCESSION NP_441990
VERSION NP_441990.1 GI:16331262
DBSOURCE REFSEQ: accession NC_000911.1
KEYWORDS .
SOURCE Synechocystis sp. PCC 6803.
ORGANISM Synechocystis sp. PCC 6803
Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE 1 (residues 1 to 714)
AUTHORS Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N.,
Sugiura,M. and Tabata,S.
TITLE Sequence analysis of the genome of the unicellular cyanobacterium

```
            Synechocystis sp. strain PCC6803. I. Sequence features in the 1
Mb
            region from map positions 64% to 92% of the genome
  JOURNAL   DNA Res. 2 (4), 153-166 (1995)
  MEDLINE   96127529
REFERENCE   2  (residues 1 to 714)
  AUTHORS   Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E.,
Nakamura,Y.,
            Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T.,
            Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
            Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
            Yamada,M., Yasuda,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular
cyanobacterium
            Synechocystis sp. strain PCC6803. II. Sequence determination of
the
            entire genome and assignment of potential protein-coding
regions
  JOURNAL   DNA Res. 3 (3), 109-136 (1996)
  MEDLINE   97061201
REFERENCE   3  (residues 1 to 714)
  AUTHORS   Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:tabata@kazusa.or.jp,
URL:http://www.kazusa.or.jp/cyano/,
            Tel:81-438-52-3933(ex.2330), Fax:81-438-52-3934)
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to
final
            NCBI review. The reference sequence was derived from BAA10060.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..714
                     /organism="Synechocystis sp. PCC 6803"
                     /db_xref="taxon:1148"
     Protein         1..714
                     /name="unknown protein"
     CDS             1..714
                     /gene="sll0169"
                     /coded_by="complement(NC_000911.1:2314780..2316924)"
                     /transl_table=11
ORIGIN
        1 mfipldfyri lgippqsgge tieqayqdrl lqlprrefsd aavtlrnqll aiayetlrdp
       61 ekrqaydqew wgamdealge alplttpele cspeqeigal lilldlgeye lvvkygepvl
      121 hdpnppaggl pqdyllsvil ahwelsrerw qqqqyefaat aslkalarlq qdndfpalea
      181 eirqelyrlr pyrilellak egqgeeqrqq glallqamvq drggiegkge dysglgnddf
      241 lkfihqlrch ltvaeqnalf lpesqrpslv asylavhslm aegvkeqdpm aiveakslii
      301 qlencqdlal ekvicelllg qtevvlaaid qgdpkivagl esklatgedp ltafytfteq
      361 wleeeivpyf rdlspetlsp kayfnnpsvq qyleqlepds fttdnsfasp allstatese
      421 tpmvhssaal pdrpltstvp srrgrsprrs rddvfpsadn ssglavttls paiaydthsl
      481 gtngiggdst sngfssnsap estskhkspr rrkkrvtikp vrfgifllcl agivggatal
      541 iinrtgdplg glledpldvf ldqpsefipd eatsrnlils qpnfnqqvgq mvvqgwldsk
      601 klafgqnydv galqsvlapn llaqqrgraq rdqaqkvyhq yehklqilay qvnpqdpnra
```

84

```
      661 tvtarveeis qpftlgnqqq kgsatkddlt vryqlvrhqg vwkidqiqvv ngpr

LOCUS       BAA10060                 714 aa            linear   BCT 04-JUL-
2001
DEFINITION  ORF_ID:sll0169~unknown protein [Synechocystis sp. PCC 6803].
ACCESSION   BAA10060
VERSION     BAA10060.1  GI:1001436
DBSOURCE    locus SYCSLRA accession D63999.1
KEYWORDS    .
SOURCE      Synechocystis sp. PCC 6803.
  ORGANISM  Synechocystis sp. PCC 6803
            Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE   1  (residues 1 to 714)
  AUTHORS   Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T.,
Miyajima,N.,
            Sugiura,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular
cyanobacterium
            Synechocystis sp. strain PCC6803. I. Sequence features in the 1
Mb
            region from map positions 64% to 92% of the genome
  JOURNAL   DNA Res. 2 (4), 153-166 (1995)
  MEDLINE   96127529
   PUBMED   8590279
REFERENCE   2  (residues 1 to 714)
  AUTHORS   Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E.,
Nakamura,Y.,
            Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T.,
            Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
            Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
            Yamada,M., Yasuda,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular
cyanobacterium
            Synechocystis sp. strain PCC6803. II. Sequence determination of
the
            entire genome and assignment of potential protein-coding
regions
  JOURNAL   DNA Res. 3 (3), 109-136 (1996)
  MEDLINE   97061201
   PUBMED   8905231
REFERENCE   3  (residues 1 to 714)
  AUTHORS   Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-1995) Satoshi Tabata, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:tabata@kazusa.or.jp,
URL:http://www.kazusa.or.jp/cyano/,
            Tel:81-438-52-3933(ex.2330), Fax:81-438-52-3934)
COMMENT     Potential protein coding regions were assigned on the basis of
            similarity search of the ORFs and GeneMark analysis.
FEATURES             Location/Qualifiers
     source          1..714
                     /organism="Synechocystis sp. PCC 6803"
                     /strain="PCC6803"
```

```
                     /db_xref="taxon:1148"
                     /note="synonym:Synechocystis PCC6803"
     Protein         1..714
                     /name="ORF_ID:sll0169
                     unknown protein"
     CDS             1..714
                     /gene="sll0169"
                     /coded_by="complement(D63999.1:47521..49665)"
                     /transl_table=11
ORIGIN
        1 mfipldfyri lgippqsgge tieqayqdrl lqlprrefsd aavtlrnqll aiayetlrdp
       61 ekrqaydqew wgamdealge alplttpele cspeqeigal lilldlgeye lvvkygepvl
      121 hdpnppaggl pqdyllsvil ahwelsrerw qqqqyefaat aslkalarlq qdndfpalea
      181 eirqelyrlr pyrilellak egqgeeqrqq glallqamvq drggiegkge dysglgnddf
      241 lkfihqlrch ltvaeqnalf lpesqrpslv asylavhslm aegvkeqdpm aiveakslii
      301 qlencqdlal ekvicelllg qtevvlaaid qgdpkivagl esklatgedp ltafytfteq
      361 wleeeivpyf rdlspetlsp kayfnnpsvq qyleqlepds fttdnsfasp allstatese
      421 tpmvhssaal pdrpltstvp srrgrsprrs rddvfpsadn ssglavttls paiaydthsl
      481 gtngiggdst sngfssnsap estskhkspr rrkkrvtikp vrfgifllcl agivggatal
      541 iinrtgdplg glledpldvf ldqpsefipd eatsrnlils qpnfnqqvgq mvvqgwldsk
      601 klafgqnydv galqsvlapn llaqqrgraq rdqaqkvyhq yehklqilay qvnpqdpnra
      661 tvtarveeis qpftlgnqqq kgsatkddlt vryqlvrhqg vwkidqiqvv ngpr
//
```

```
LOCUS       AY074283                2857 bp    mRNA    linear   PLN 26-APR-
2002
DEFINITION  Arabidopsis thaliana unknown protein (At3g19180) mRNA, complete
            cds.
ACCESSION   AY074283
VERSION     AY074283.1  GI:18377659
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2857)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G.,
Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C.,
Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2857)
  AUTHORS   Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C.,
Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M.,
Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2002) Plant Gene Expression Center, 800
Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M.,
Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out
the
            sequencing and annotation of the RAFL cDNAs: Yamada,K.,
Banh,J.,
            Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
            Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
```

```
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed
equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A.
(SSP/PGEC)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1..2857
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="RAFL09-57-L03 (R19126)"
                     /note="This clone is in a modified pBluescript vector
                     (FLC-1) as a BamHI/XhoI insert.
                     ecotype: Columbia"
     gene            1..2857
                     /gene="At3g19180"
     5'UTR           1..134
                     /gene="At3g19180"
     CDS             135..2594
                     /gene="At3g19180"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAL66980.1"
                     /db_xref="GI:18377660"

/translation="MPVAYTFPVLPSSCLLCGISNRSTSFVVDRPELQISGLLVVRSE

SGEFFGSGLSLRRFQREGRRRLNAAGGGIHVVDNAPSRTSSLAASTSTIELPVTCYQL

IGVSEQAEKDEVVKSVINLKKTDAEEGYTMEAAAARQDLLMDVRDKLLFESEYAGNLK

EKIAPKSPLRIPWAWLPGALCLLQEVGQEKLVLDIGRAALRNLDSKPYIHDIFLSMAL

AECAIAKAAFEVNKVSQGFEALARAQSFLKSKVTLGKLALLTQIEESLEGLAPPCTLD

LLGLPRTPENAERRRGAIAALRELLRQGLSVEASCQIQDWPCFLSQAISRLLATEIVD

LLPWDDLAITRKNKKSLESHNQRVVIDFNCFYMVLLGHIAVGFSGKQNETINKAKTIC

ECLIASEGVDLKFEEAFCSFLLKQGSEAEALEKLKQLESNSDSAVRNSILGKESRSTS

ATPSLEAWLMESVLANFPDTRGCSPSLANFFRAEKKYPENKKMGSPSIMNHKTNQRPL

STTQFVNSSQHLYTAVEQLTPTDLQSPVVSAKNNDETSASMPSVQLKRNLGVHKNKIW

DEWLSQSSLIGRVSVVALLGCTVFFSLKLSGIRSGRLQSMPISVSARPHSESDSFLWK

TESGNFRKNLDSVNRNGIVGNIKVLIDMLKMHCGEHPDALYLKSSGQSATSLSHSASE

LHKRPMDTEEAEELVRQWENVKAEALGPTHQVYSLSEVLDESMLVQWQTLAQTAEAKS
```

```
CYWRFVLLHLEVLQAHIFEDGIAGEAAEIEALLEEAAELVDESQPKNAKYYSTYKIRY
                     ILKKQEDGLWKFCQSDIQIQK"
     misc_difference 937
                     /gene="At3g19180"
                     /note="compared to genomic sequence resulting in an
amino
                     acid sequence difference"
                     /replace="a"
     3'UTR           2595..2857
                     /gene="At3g19180"
     misc_difference 2841
                     /gene="At3g19180"
                     /note="not present in genomic sequence"
BASE COUNT      808 a    584 c    644 g    821 t
ORIGIN
        1 actgtcaaaa ctcaaaagcc ttgagaccaa atttccgatt ttttctcctc tgaagaaatc
       61 caacaaattg taccatgatt ccagcttcac tctacttctt ctagggttcg ttcgttttct
      121 ggagctgttg cgcaatgcca gtagcttaca catttccagt tctcccttct tcttgtctgc
      181 tttgcggaat ctccaatcgc agcaccagct tcgtcgtaga tcgcccggag cttcagatct
      241 caggtctcct cgtcgttcgt tctgaatccg gtgaattctt cggttctggt ttatctttgc
      301 ggcggtttca gcgagaagga cggaggaggt tgaatgctgc tggtggtggt atccatgtcg
      361 tcgacaatgc gccgtctcgt acttcttctc tcgctgcatc tacctctaca atcgaactcc
      421 cggttacgtg ttaccagctt atcggagttt ctgagcaagc tgagaaagac gaggtcgtta
      481 agtcggttat aaatttgaaa aaaactgatg ctgaagaggg ttatacaatg gaagctgctg
      541 cagctcgcca ggatcttctc atggatgtta gggataaact tctttttgaa tcagaatatg
      601 ctggtaacct aaaagaaaag attgctccta aatctcctct cagaattccg tgggcatggt
      661 tgcctggtgc tctatgcctt cttcaagagg ttggacaaga aaaacttgtg ctggatattg
      721 gccgggctgc tctcaggaac cttgattcaa agccatatat tcatgatata ttcttatcta
      781 tggcacttgc tgagtgtgca attgccaagg ctgctttcga ggttaacaag gtctctcaag
      841 gatttgaagc tcttgctcgt gctcaaagtt ttctgaagag taaagttact cttgggaaac
      901 ttgcattgtt aactcagatt gaggagtcac tagaggggct tgcaccacct tgcacattgg
      961 atctactggg cctgccacgc acgccagaaa atgcagagag gaggcgaggt gcaattgccg
     1021 cgctacgcga actgctcaga cagggcctta gtgttgaagc ttcatgtcaa attcaagact
     1081 ggccatgctt tttgagccag gcaattagca ggttattggc cacagagatt gtcgatcttc
     1141 ttccatggga tgatttagcc attacacgga aaaataaaaa atcactggaa tcccacaatc
     1201 aaagagttgt tattgatttt aattgtttct acatggtgtt acttggtcac atcgctgttg
     1261 gattttcagg caagcaaaat gaaacgatta ataaagcaaa aacgatatgc gaatgtctca
     1321 tagcatcaga aggtgttgat ctgaaatttg aggaagcttt ttgctcattt cttctaaaac
     1381 agggttccga ggcagaggcc ctggaaaaac ttaagcagct ggaatcaaat tcagactctg
     1441 ccgttcgtaa ttcgatcttg gggaaagagt cgagaagtac ttctgctact ccctcactgg
     1501 aagcgtggct aatggagtcc gtgcttgcta actttccaga cacaaggggt tgttctccat
     1561 ctttggccaa ttttttccgg gctgaaaaga aatatccaga aaacaagaaa atggggtcac
     1621 cttcgatcat gaatcataag acgaaccaaa gaccactttc cacaacacag ttcgtgaact
     1681 cgtcacaaca tctttataca gctgtcgagc agttgacacc aacagatttg cagagcccag
     1741 tggtatcagc caagaataat gatgaaacca gtgccagtat gccatctgtt caactgaaga
     1801 ggaaccttgg tgtacacaaa aataaaatat gggatgagtg gctctctcaa agcagtttga
     1861 tcggaagggt atctgttgtt gctttactgg gttgcaccgt gttcttctct ctgaagctat
     1921 caggcattag gtctggtaga ctacagagta tgcctatatc ggttttctgct aggccgcatt
     1981 cagaatcaga ttctttttctg tggaaaacag agtctgggaa tttcagaaaa aaccttgatt
     2041 ctgtgaatag aaatggtatc gtgggaaaca tcaaagtgct cattgacatg ttaaagatgc
     2101 attgtggcga acatccggat gccctgtatc tgaaaagctc tggtcaatca gctacatcat
     2161 tgtctcattc tgcgtcagaa ctgcataaga gaccaatgga tacagaagaa gcggaagagc
     2221 ttgtgagaca gtgggaaaat gttaaggctg aagctcttgg accaacacat caagtttata
     2281 gcctttccga agtccttgat gaatccatgc ttgtccagtg gcaaacattg gcacaaacag
```

```
2341 cagaggcgaa atcctgttat tggaggttcg ttctgcttca tcttgaggtt ttgcaagcac
2401 atatattcga agatggtatt gctggtgagg ctgcagaaat cgaagctctt ctggaggaag
2461 cagcagaatt agttgatgaa tctcagccca aaaacgcaaa atattatagc acttacaaga
2521 tccgatatat tctgaagaag caagaagatg gattgtggaa attctgccaa agcgatattc
2581 aaatacagaa gtgaaaatcc cccagaaaaa aaagctcatc atctaactaa aggttgtagc
2641 atcaacagta gaacatggga tcatttagct aacggttgtt cttgtttacc taacggtgta
2701 ggaaagtctc aggtttgttt ctttattcct tagtaaccca caggatttgt ctttgtagat
2761 tcttttgatt tcaatgtgtt tatggataaa caaacttctt gagtattttt tttattatta
2821 ttgtaaagcg ttactgatca caaaaaaaaa aaaaaaa
//
```

```
LOCUS       AAL66980                819 aa            linear   PLN 26-APR-
2002
DEFINITION  unknown protein [Arabidopsis thaliana].
ACCESSION   AAL66980
VERSION     AAL66980.1  GI:18377660
DBSOURCE    accession AY074283.1
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (residues 1 to 819)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G.,
Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C.,
Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 819)
  AUTHORS   Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C.,
Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M.,
Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2002) Plant Gene Expression Center, 800
Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M.,
Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out
the
            sequencing and annotation of the RAFL cDNAs: Yamada,K.,
Banh,J.,
            Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
            Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
```

```
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed
equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A.
(SSP/PGEC)
            contributed equally to this work as PIs.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..819
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="RAFL09-57-L03 (R19126)"
                     /note="This clone is in a modified pBluescript vector
                     (FLC-1) as a BamHI/XhoI insert.
                     ecotype: Columbia"
     Protein         1..819
                     /product="unknown protein"
     CDS             1..819
                     /gene="At3g19180"
                     /coded_by="AY074283.1:135..2594"
ORIGIN
        1 mpvaytfpvl psscllcgis nrstsfvvdr pelqisgllv vrsesgeffg sglslrrfqr
       61 egrrrlnaag ggihvvdnap srtsslaast stielpvtcy qligvseqae kdevvksvin
      121 lkktdaeegy tmeaaaarqd llmdvrdkll feseyagnlk ekiapksplr ipwawlpgal
      181 cllqevgqek lvldigraal rnldskpyih diflsmalae caiakaafev nkvsqgfeal
      241 araqsflksk vtlgklallt qieeslegla ppctldllgl prtpenaerr rgaiaalrel
      301 lrqglsveas cqiqdwpcfl sqaisrllat eivdllpwdd laitrknkks leshnqrvvi
      361 dfncfymvll ghiavgfsgk qnetinkakt icecliaseg vdlkfeeafc sfllkqgsea
      421 ealeklkqle snsdsavrns ilgkesrsts atpsleawlm esvlanfpdt rgcspslanf
      481 fraekkypen kkmgspsimn hktnqrplst tqfvnssqhl ytaveqltpt dlqspvvsak
      541 nndetsasmp svqlkrnlgv hknkiwdewl sqssligrvs vvallgctvf fslklsgirs
      601 grlqsmpisv sarphsesds flwktesgnf rknldsvnrn givgnikvli dmlkmhcgeh
      661 pdalylkssg qsatslshsa selhkrpmdt eeaeelvrqw envkaealgp thqvyslsev
      721 ldesmlvqwq tlaqtaeaks cywrfvllhl evlqahifed giageaaeie alleeaaelv
      781 desqpknaky ystykiryil kkqedglwkf cqsdiqiqk
```

```
LOCUS       NC_003074           23465812 bp    DNA     linear   PLN 10-JAN-
2002
DEFINITION  Arabidopsis thaliana chromosome 3, complete sequence.
ACCESSION   NC_003074
VERSION     NC_003074.2  GI:18426881
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 23465812)
  AUTHORS   Town,C.D., Haas,B.J., Wu,D., Maiti,R., Hannick,L.I., Chan,A.P.,
            Tallon,L.J., Rooney,T., Utterback,T.R., VanAken,S.E.,
            Feldblyum,T.V., White,O. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 3 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 23465812)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) The Institute for Genomic Research,
9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to
final
            NCBI review. The reference sequence was derived from AE102093.
            On Jan 30, 2002 this sequence version replaced gi:15228160.
            Address all correspondence to:at@tigr.org

             Genes were identified by a combination of several methods:
Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark
Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a
variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named
to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without
protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-
SE/).
```

```
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1..23465812
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
...
...
gene            6632806..6639031
                     /gene="At3g19180"
                     /note="MVI11.9; predicted by genscan+"
     mRNA            join(<6632806..6633108,6633408..6633521,
                     6633599..6633736,6633812..6633916,6634008..6634130,
                     6634812..6634907,6635016..6635168,6635577..6635642,
                     6635728..6636480,6636588..6636778,6636865..6636945,
                     6637595..6637697,6637777..6637843,6638047..6638104,
                     6638203..6638365,6638457..6638663,6638749..6638929,
                     6639021..>6639031)
                     /gene="At3g19180"
                     /transcript_id="NM_112805.1"
                     /db_xref="GI:18402148"
     CDS
join(6632806..6633108,6633408..6633521,6633599..6633736,
                     6633812..6633916,6634008..6634130,6634812..6634907,
                     6635016..6635168,6635577..6635642,6635728..6636480,
                     6636588..6636778,6636865..6636945,6637595..6637697,
                     6637777..6637843,6638047..6638104,6638203..6638365,
                     6638457..6638663,6638749..6638929,6639021..6639031)
                     /gene="At3g19180"
                     /codon_start=1
                     /protein_id="NP_188549.1"
                     /db_xref="GI:15230315"
```

Second Set

dbEST Id: 12028705
EST name: BJ258222
GenBank Acc: BJ258222
GenBank gi: 20081080

CLONE INFO
Clone Id: whh6h02 (5')
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GGCCGTCGGCAAATACTGCAGNTTGCACATGATACTCTCACAAACCAGAGCTCCCGCACC

GAGTATGACCGCGCGCTCTCTGAGGACCGTGACGCGGCGCTCACACTGGATGTTGCTTGG

GACAAGGTTCCGGGTGTGCTATGTGCCCTTCAGGAGGCTGGGGAGGCACAGGCAGTGCTT

GCAATTGGAGAGCACTTACTGGAGGACCGCCCGCCCAAGCGGTTCAAGCAGGATGTGGTG

CTGGCAATGGCGCTCGCTTATGTGGACATATCAAGGGATGCAATGGCGGCTAGCCCTCCA

GATGTAATCCGCTGCTGTGAGGTGCTTGAAAGGGCTCTCAAGCTCTTGCAGGAGGATGGG

GCAATCAACCTTGCACCTGGTCTGCTTTCACAAATTGATGAAACTCTGGAGGAGATCACA

CCTCGTTGTGTTTTGGAGCTTCTTGCCCTTNCTCTTGATGAAAAACATCANATTGAACGC
CANNAANGNNT

Entry Created: Apr 8 2002
Last Updated: Apr 8 2002

LIBRARY
Lib Name: Y. Ogihara unpublished cDNA library, Wh_h
Organism: Triticum aestivum
Cultivar: Chinese Spring
Tissue type: spike at heading date
Develop. stage: Feekes' scale 10.5

SUBMITTER
Name: Tadasu Shin-i
Lab: Center For Genetic Resource Information
Institution: National Institute of Genetics
Address: 1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
E-mail: tshini@genes.nig.ac.jp

CITATIONS
Title: Expressed genes in Triticum aestivum.
Authors: Ogihara,Y., Murai,K.

Year: 2002
Status: Unpublished

dbEST Id: 12455031
EST name: GA__Ed0029A07f
GenBank Acc: BQ410206
GenBank gi: 21097893
CLONE INFO
Clone Id: GA__Ed0029A07f
Source: CUGI
DNA type: cDNA
PRIMERS
Sequencing: TAATACGACTCACTATAGGG
PolyA Tail: Unknown

SEQUENCE

AATTGCAGAAGGCATTGTTCGCAAGTGGCAGAACATTAAATCTGAGGCGTTTGGACCTGA

TCACCGCCTTGATAAATTGCCAGAGGTTCTGGATGGTCAAATGTTGAAGACATGGACAGA

TCGTGCAGCCGAAATCGCTCAGCTTGGTTGGGTATATGAATATAGTCTACTGAACATGGC

CATTGACAGTGTTACCCTTTCACTAGATGGCCAGCGAGCTGTAGTCGAAGCTACTCTGGA

AGAATCCACCTGCTTGACTGATGTTCATCATCCGGAGAACAATGCCTCTAATGTAAACTC

CTACACCACGAGATATGAGATGTCTTGTTCCAACTCAGGCTGGAAAATCACTGAAGGATC

TGTCTACAAATCTTAACTATGATGTATAAAGCATAAAAAGCCTGAAAGCTCCAATGTGGT

TACCAGCTTTGCCTTTTTACGTAGCTATATTTGTTATATTGTTTGAGAAAACAAGAGTTA

GCGTTTTCCAGTCATGCAAGCAGTTCAAATTAAAAGAGGCAATGCTTNTCATGGANAACN
AAATG

Quality: High quality sequence stops at base: 538
Entry Created: May 22 2002
Last Updated: May 22 2002
COMMENTS
Total High Quality bases = 521
LIBRARY
Lib Name: Gossypium arboreum 7-10 dpa fiber library
Organism: Gossypium arboreum
Strain: AKA
Cultivar: 8400
Tissue type: Fibers isolated from bolls harvested 7-10 dpa
Lab host: E. coli
Vector: pBK-CMV
R. Site 1: EcoRI
R. Site 2: XhoI

SUBMITTER
Name: Wing RA
Lab: Clemson University Genomics Institute
Institution: Clemson University
Address: 100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293

E-mail: rwing@clemson.edu

CITATIONS

Title: An integrated analysis of the genetics, development, and evolution of the cotton fiber

Authors: Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry,D., Wood,T.C., Leslie,A., Wilkins,T.A.

Year: 2000

Status: Unpublished

dbEST Id: 12551917
EST name: AJ485537
GenBank Acc: AJ485537
GenBank gi: 21201492

CLONE INFO
Clone Id: S0001100068E09F1
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GATGAGCCCATACAGATTCCTAAAATGGATGCGAAGCTGGCAGAAGATATTGTTCGCAAG

TGGCAGAGCATCAAATCCAAGGCCTTGGGATCAGATCATTCTGTTGCATCATTGCAAGAG

GTTCTTGATGGCAACATGCTGAAGGTATGGACAGACCGAGCAGCAGAGATTGAGCGCAAA

GGCTGGTTCTGGGACTACACGCTGTTCAACGTGGCGATCGACAGCATCACCGTCTCCCTG

GACGGACGGCGGGCGACCGTGGAGGCGACAATTGAGGAGGCGGGTCAGCTCACCGACGCA

ACCGACCCCAGGAACGATGATTTGTACGACACTAAGTACACCACCCGGTACGAGATGGCC

TTCACCGGACCAGGAGGGTGGAAGATAACCGAAGGCGCAGTCCTCAAGTCGTCATAGGGC

Entry Created: May 24 2002
Last Updated: May 24 2002

LIBRARY
Lib Name: S00011
Organism: Hordeum vulgare
Develop. stage: Developing seed
Description: 12,15,18 days after pollination

SUBMITTER
Name: Schulman AH
Lab: Institute of Biotechnology
Institution: University of Helsinki
Address: P.O.Box 56 (Viikinkaari 6A), University of Helsinki
FIN-00014, Finland

CITATIONS
Title: Barley EST's
Authors: Saren,A.-M., Tanskanen,J., Paulin,L., Schulman,A.H.
Year: 2002
Status: Unpublished

dbEST Id: 12032032
EST name: BJ263824
GenBank Acc: BJ263824
GenBank gi: 20084407

CLONE INFO
Clone Id: whh6h02 (3')
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

CTGCAAATCTAGCACTATGTTTCTCTTTATCTCCAGGATCTAGCCTAGCACCAACAATCC

AAATACAACACAAGAAAAATAAAGCTCTTCGTCGATCACATCAGACTAACGCAACTATCG

GTCTTCCAAACTAAAAAGGGCCTAGACTGCCTGCTTATTTACACACCCCCAAAAGAAAAC

TGGAAGGAATTAACAAACTTAATGAGGTTACCGCACACCAACTACCCTAAGACGACTTGA

GGACCGCGCCTTCCATTATCTTCCACCCTCCTAGTCCGGTGAAGGTCATCTCATACCGGG

TGGTGTACTTCGTGTCGTACGAGTCGTTGTTCTTGGGGTCGGTTGCGTCGATGAGCTGGC

CTGCCTCCTCGATCGTTGCCTCCACGGTCGCCCGCCGTCCGTCCAGGGAGACCGTGATGC

TGTCGATCGCCACGTCAGACAGTGTGTAGTCCCAGAACCAGCCTTTGCGCCCGATCTCCG

CTGCTCGGTCCGTCCATACCTTCAGCATGTTGCCATCAAGAACCTCTTGCAATGATTCCA

CAGAATGATCTGATCCCAAGGCCTTGGTTTTGATACTCTGCCACTTGCGAACAATATCTT
CTGCCA

Entry Created: Apr 8 2002
Last Updated: Apr 8 2002

LIBRARY
Lib Name: Y. Ogihara unpublished cDNA library, Wh_h
Organism: Triticum aestivum
Cultivar: Chinese Spring
Tissue type: spike at heading date
Develop. stage: Feekes' scale 10.5

SUBMITTER
Name: Tadasu Shin-i
Lab: Center For Genetic Resource Information
Institution: National Institute of Genetics
Address: 1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
E-mail: tshini@genes.nig.ac.jp

CITATIONS

Title: Expressed genes in Triticum aestivum.
Authors: Ogihara,Y., Murai,K.
Year: 2002
Status: Unpublished

dbEST Id: 12455032
EST name: GA__Ed0029A07r
GenBank Acc: BQ410207
GenBank gi: 21097894

CLONE INFO
Clone Id: GA__Ed0029A07r
Source: CUGI
DNA type: cDNA

PRIMERS
Sequencing: TAATACGACTCACTATAGGG
PolyA Tail: Unknown

SEQUENCE

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTTGCCTCTTTTAATTTGAA

CTGCTTGCCTGACTGGAAAACCCTAACTCTTGTTTTCTCAAACAATTTAACAAATATAGC

TCCCTAAAAAGGCAAAGCTGGTAACCACATTGGAGCTTTCAGGCTTTTTATGCTTTATAC

ATCATAGTTAAAATTTGTAGACAGATCCTTCAGTGATTTTCCAACCTGAGTTGGAACAAA

ACATCTCATATTTCGTGGGGTAGGAGTTTACATTACAGGCATTGTTCTCCGGATGATGAA

CATTACTCAAGCCGGGGGGTTCTTCCAAAATAACTTCGACTACAGCTCGCTGGCCATTTA

ATGAAAGGGTAACACTGTCAATGGCCCTGTTCAGTCAACTTTATTCATATACCCAACCCA

GCTGACCGATTTCGGCTGCACCAACTGTCCATGTTTTCAACATTTGACCATCCAAAACCT

TTGGCAATTTATCAAGGGGGGGATCAAGTCCAAACGCCTCAGATTTAATGTTCTGCCACT
TGCGAACAATGCCTTTTGCAATT

Quality: High quality sequence starts at base: 3
Quality: High quality sequence stops at base: 554

Entry Created: May 22 2002
Last Updated: May 22 2002

COMMENTS
Total High Quality bases = 222

LIBRARY
Lib Name: Gossypium arboreum 7-10 dpa fiber library
Organism: Gossypium arboreum
Strain: AKA
Cultivar: 8400
Tissue type: Fibers isolated from bolls harvested 7-10 dpa
Lab host: E. coli
Vector: pBK-CMV
R. Site 1: EcoRI
R. Site 2: XhoI

SUBMITTER

Name: Wing RA
Lab: Clemson University Genomics Institute
Institution: Clemson University
Address: 100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
E-mail: rwing@clemson.edu

CITATIONS
Title: An integrated analysis of the genetics, development, and evolution of the cotton fiber
Authors: Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry,D., Wood,T.C., Leslie,A., Wilkins,T.A.
Year: 2000
Status: Unpublished

dbEST Id: 12551919
EST name: AJ485539
GenBank Acc: AJ485539
GenBank gi: 21201494

CLONE INFO
Clone Id: S0001100117E11F1
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GATGAGCCCATACAGATTCCTAAAATGGATGCGAAGCTGGCAGAAGATATTGTTCGCAAG

TGGCAGAGCATCAAATCCAAGGCCTTGGGATCAGATCATTCTGTTGCATCATTGCAAGAG

GTTCTTGATGGCAACATGCTGAAGGTATGGACAGACCGAGCAGCAGAGATTGAGCGCAAA

GGCTGGTTCTGGGACTACACGCTGTTCAACGTGGCGATCGACAGCATCACCGTCTCCCTG

GACGGACGGCGGGCGACCGTGGAGGCGACAATTGAGGAGGCGGGTCAGCTCACCGACGCA

ACCGACCCCAGGAACGATGATTTGTACGACACTAAGTACACCACCCGGTACGAGATGGCC

Entry Created: May 24 2002
Last Updated: May 24 2002

LIBRARY
Lib Name: S00011
Organism: Hordeum vulgare
Develop. stage: Developing seed
Description: 12,15,18 days after pollination

SUBMITTER
Name: Schulman AH
Lab: Institute of Biotechnology
Institution: University of Helsinki
Address: P.O.Box 56 (Viikinkaari 6A), University of Helsinki
FIN-00014, Finland

CITATIONS
Title: Barley EST's
Authors: Saren,A.-M., Tanskanen,J., Paulin,L., Schulman,A.H.
Year: 2002
Status: Unpublished

dbEST Id: 12426231
EST name: AJ463103
GenBank Acc: AJ463103
GenBank gi: 21062023

CLONE INFO
Clone Id: S0000200015A03F1
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

TGATGGCAACATGCTGAAGGTATGGACAGACCGAGCAGCAGAGATTGAGCGCAAAGGCTG

GTTCTGGGACTACACGCTGTTCAACGTGGCGATCGACAGCATCACCGTCTCCCTGGACGG

ACGGCGGGCGACCGTGGAGGCGACAATTGAGGAGGCGGGTCAGCTCACCGACGCAACCGA

CCCCAGGAACGATGATTTGTACGACACTAAGTACACCACCCGGTACGAGATGGCCTTCAC

CGGACCAGGAGGGTGGAAGATAACCGAAGGCGCAGTCCTCAAGTCGTCATAGGGCGTTCA

Entry Created: May 21 2002
Last Updated: May 24 2002

LIBRARY
Lib Name: S00002
Organism: Hordeum vulgare
Cultivar: Saana
Develop. stage: Embryo
Description: 1 day after pollination

SUBMITTER
Name: Schulman AH
Lab: Institute of Biotechnology
Institution: University of Helsinki
Address: P.O.Box 56 (Viikinkaari 6A), University of Helsinki
FIN-00014, Finland

CITATIONS
Title: Barley EST's
Authors: Saren,A.-M., Tanskanen,J., Paulin,L., Schulman,A.H.
Year: 2002
Status: Unpublished

dbEST Id: 12172134
EST name: WHE2493_E05_J09ZT
GenBank Acc: BQ169059
GenBank gi: 20315019

CLONE INFO
Clone Id: WHE2493_E05_J09
DNA type: cDNA

PRIMERS
Sequencing: T7 primer
PolyA Tail: Unknown

SEQUENCE

TTTTTTTTTTTTTTTTTTTTTTTTTTTCAGCGGCAAATTCAGCACTATGTTTCTCTTAT

CCCCAACTCAAAGATCTTCTAAGCTAGCAATAATCCGAAAACGACACAGGGAAAAACAAA

GCTCATCGCTGATTGCACATCAGACTAACCAAACTATCTCCAACTTCCAAACTGAGAAGG

GCCTAGACTGCTTATTTACACACCAAAAAGAACACGGGAGGAATCAATCAACAAAGGTCT

ACTGCACACCGAACGCCCTATGACGACTTGAGGACCGCACCTTCTGTTATCTTCCACCCT

CCTGGTCCAGTGAAGGTCATCTCGTACCGGGTGGTGTACTTAGTGTCGTACAAATCGTTG

TTCCTGGGGTCGGTTGCATCGGTAAGCTGGCCTGCCTCCTCAATTGTCGCCTCCACAGTC

GCCCGTCGTCCGTCCAGGGAGACGGTGATGCTGTCAATCGCCACGTCGGACAGCGTGTAG

TCCCAGAACCAGCCTTTGCGCTCGATCTCTGCTGCTCGGTCCCTCCATACCTTCAGCATG
TTGCCATCA

Entry Created: Apr 25 2002
Last Updated: Apr 25 2002

COMMENTS
This EST was generated by sequencing from the 3' end of the clone. Sequences have been trimmed to remove vector sequence
and low quality sequence with phred score less than 20.

LIBRARY
Lib Name: Triticum monococcum early reproductive apex cDNA library
Organism: Triticum monococcum
Cultivar: DV92
Tissue type: Early reproductive apex
Develop. stage: Seven week-old plants
Lab host: E. coli XLOLR
Vector: Lambda Uni-ZAP XR, excised phagemid
R. Site 1: EcoRI
R. Site 2: XhoI
Description: The tissue, total RNA, and poly(A) RNA were prepared from

apex at double-ridge stage to terminal-spikelet stage during transition from vegetative state to flower state, a cDNA library was made, and the cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova, J. Dubcovsky). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).

SUBMITTER
Name: Olin Anderson
Institution: US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
Address: 800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
E-mail: oandersn@pw.usda.gov

CITATIONS
Title: The structure and function of the expressed portion of the wheat genomes - Early reproductive apex cDNA library from Triticum monococcum
Authors: Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Stamova,B., Tong,J.C.
Year: 2001
Status: Unpublished

dbEST Id: 12506802
EST name: BJ482132
GenBank Acc: BJ482132
GenBank gi: 21160594

CLONE INFO
Clone Id: bah63k10 (5')
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GCGAGNAAGGACGAGNATCGTCAAGTCGGCCATCGAGCTGAGGAAATCGGAGATCGAAGA

TGGGTACACGGAGGAGGTGTCCACCTGCAGACAGGCTCTGCTGCTGGACGTGAGAGACAA

GCTTCTCTTTGAACAGGAGTACGCAGGAAGCACCAGGGCCAAGGTTCCGCCCAGATCCTC

TCTTCATATACCCTGGAGCTGGTTGCCTGCTGCCTTGTGTGTCTTGCAGGAGGTTGGGGA

AGAGAAGCTGGTCTTGGACATTGGTCAGGCAGCTCTACGACGCCCTGATTCTAAGCCATA

TGCTCACGATGTACTTCTTGCAATGGCACTAGCTGAATGCTCCATTGCAAAAGCTAGCTT

TGAAAAAAGTAAAGTATCTCTTGGCTTTGAGGCTCTAGCACGTGCTCAATATCTTTTGAG

GAAAAAACCATCTTTAGAGAAGATGCCTCTTCTTGAGCAGATCGAAGAATCACTTGAAGA

GCTTGCACCAGCTTGCACTCTAGAGGTTTTAAGCCTGCCCCGTACACCTGAAAATTCTGA
ACGCAGGCGTGGTGCTATTGCAGCTCTCTGTGA

Entry Created: May 23 2002
Last Updated: May 23 2002

LIBRARY
Lib Name: K. Sato unpublished cDNA library, strain H602 adult, heading
stage top three leaves
Organism: Hordeum vulgare subsp. spontaneum
Strain: H602
Tissue type: top three leaves
Develop. stage: adult, heading stage

SUBMITTER
Name: Tadasu Shin-i
Lab: Center For Genetic Resource Information
Institution: National Institute of Genetics
Address: 1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
E-mail: tshini@genes.nig.ac.jp

CITATIONS

Title: Barley EST sequencing project in NIG and Okayama Univ
Authors: Sato,K., Saisho,D., Takeda,K.
Year: 2002
Status: Unpublished

dbEST Id: 12601756
EST name: 27-E011788-006-050-F04-T3
GenBank Acc: BQ490457
GenBank gi: 21335077

CLONE INFO
Clone Id: F-4-50
DNA type: cDNA

PRIMERS
Sequencing: T3 'AATTAACCCTCACTAAAGGG'
PolyA Tail: Unknown

SEQUENCE

GCATAACACGGCAAGAAGATGTTGCAGTTAATGGCTTTGGAAATGAGGATGTTACAATGG

AGCTTGGCCGTGATAACACTTTAGATTATGTGAATTTAGCCAGTTCAAATTTTACTGAAG

ATAATATCGAGCAAGAATCGGTTACTGAGAAGATAAAAGATTTAGGTGTGAAGGTTATGT

GTGCCGGTGTGGTGATTGGACTGACAACTTTGGCTGGCATGAAACTTTTGCCTGGCAGAA

GTGGGTCTGCCATTCCACACAGGCATCTTGGTTCTGCTGTGGCTTCTGATGTCTCCAGTG

TGGGGCTCTCAGTAAATGAAACTACTGAGGAGAAAGTACCAAAAATGGATGCAAGACTTG
CAGAAGTTCTAGTTAGAAGATGGCAGAACGTTAAATCACA
Quality: High quality sequence stops at base: 400

Entry Created: Jun 7 2002
Last Updated: Jun 7 2002

LIBRARY
Lib Name: Sugar beet MPIZ-ADIS-006 Lambda Zap II library
Organism: Beta vulgaris
Organ: shoot and root
Develop. stage: 4 week old pot-grown plants
Vector: pBluescript SK- from lambda ZAP II
Description: cDNA (lambda ZAP-II) library from sugar beet, whole plant
mRNA, Prepared using the Stratagene UniZAP cDNA kit,
cloning
sites EcoRI-XhoI, primer sites and orientation:
rev-T3-SacI-SK-EcoRI-GGCACGAGG-5pr-cDNA-polyA-XhoI-KpnI-T7-
u
ni

SUBMITTER
Name: Weisshaar B
Lab: ADIS DNA core facility at MPIZ
Institution: Max-Planck-Institute for Plant Breeding Research
Address: Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
E-mail: weisshaa@mpiz-koeln.mpg.de

CITATIONS

Title: EST sequencing, annotation and macroarray expression analysis of more than 3000 sugar beet cDNAs identifies genes with root-specific expression pattern.
Authors: Bellin,D., Werber,M., Theis,T., Weisshaar,B., Schneider,K.
Year: 2002
Status: Unpublished

```
>gi|22486832|gb|BU046755.1|BU046755   PP_LEa0027I04f Peach developing fruit
mesocarp Prunus persica cDNA
           clone PP_LEa0027I04f.
          Length = 631

 Score =  256 bits (653), Expect = 7e-67
 Identities = 132/198 (66%), Positives = 149/198 (75%), Gaps = 4/198 (2%)
 Frame = +1

Query: 315 REKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVYEVALALVAQAFIGKKPHLXXXXXX 374
           RE FMNEAFL MTAAEQVDLFVATPSNIPAESFEVY VALALVAQAF+GKKPH
Sbjct: 31  RENFMNEAFLHMTAAEQVDLFVATPSNIPAESFEVYGVALALVAQAFVGKKPHHIQDAEN 210

Query: 375 XXXXXXXXXVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDECRMWLGLDSEDSQY 434
                  V A+       Y T+ + EIDF LERGLC+LL+G +D+ R WLGLDS DS Y
Sbjct: 211 LFQKLQQSKVTAVGHSLDNYITKESSEIDFALERGLCSLLLGDLDDSRSWLGLDSNDSPY 390

Query: 435 RNPAIVEFVLENSNRDDNDD----LPGLCKLLETWLAGVVFPRFRDTKDKKFKLGDYYDD 490
           RNP++V+FVLENS  DD++D    LPGLCKLLETWL  VVFPRFRDTKD +F+LGDYYDD
Sbjct: 391 RNPSVVDFVLENSKDDDDNDNDNDLPGLCKLLETWLMEVVFPRFRDTKDIEFRLGDYYDD 570

Query: 491 PMVLSYLERVEVVQGSPL 508
           P VL YLER++  GSPL
Sbjct: 571 PTVLRYLERLDGTNGSPL 624

LOCUS       BU046755                631 bp    mRNA    linear   EST 26-AUG-
2002
DEFINITION  PP_LEa0027I04f Peach developing fruit mesocarp Prunus persica
cDNA
            clone PP_LEa0027I04f, mRNA sequence.
ACCESSION   BU046755
VERSION     BU046755.1  GI:22486832
KEYWORDS    EST.
SOURCE      Prunus persica (peach)
ORGANISM    Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
            Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons;
            core eudicots; rosids
            ; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Callahan,A., Palmer,M., Main,D., Wing,R. and Abbott,A.
  TITLE     Peach Model Genome for Rosaceae
  JOURNAL   Unpublished
COMMENT     Contact: Abbott, A.
            Dept of Genetics and Biochemistry
            Clemson University
            122 Long Hall, Clemson University, Clemson, SC 29634, USA
            Tel: 864 656 3060
            Fax: 864 656 6879
            Email: aalbert@clemson.edu
            Total High Quality bases = 523
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 631.
FEATURES             Location/Qualifiers
     source          1..631
```

```
                     /organism="Prunus persica"
                     /mol_type="mRNA"
                     /cultivar="Loring"
                     /db_xref="taxon:3760"
                     /clone="PP_LEa0027I04f"
                     /tissue_type="Mesocarp"
                     /lab_host="E. coli"
                     /clone_lib="Peach developing fruit mesocarp"
                     /note="Vector: pBluescript II SK(-); Site_1: EcoRI;
                     Site_2: XhoI; authority=Prunus persica L. Batsh; The
                     sequence has been trimmed to remove vector sequence
and
                     contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis go to
                     http://www.genome.clemson.edu/projects/peach. To order
                     this clone go to http://www.genome.clemson.edu/orders"
BASE COUNT      174 a    123 c    155 g    178 t      1 others
ORIGIN
        1 gcagttgcaa ttgctggggg ngattcacta cgtgaaaatt tcatgaacga ggccttcttg
       61 catatgactg cagctgagca ggttgattta tttgtagcta cccccagtaa tatcccggca
      121 gaaagctttg aagtttatgg ggtggctctt gcgcttgttg ctcaagcctt tgttggtaaa
      181 aaacctcatc acattcaaga tgctgaaaac ctattccaga aacttcagca gtctaaggta
      241 acagctgtag gacattctct tgacaactat ataaccaaag aaagcagtga gatagacttt
      301 gctttggaga ggggactctg ttcacttctt ctaggggacc ttgatgacag tcgttcgtgg
      361 ttgggcctag acagtaatga ttcaccatat agaaatccat ctgttgtaga ctttgtcttg
      421 gagaactcaa aggatgacga tgacaatgac aatgacaatg atcttcctgg actttgcaag
      481 ctattggaga cgtggttgat ggaggtggta ttccccaggt ttagagacac caaagacata
      541 gagttcagac tgggagacta ctatgatgat cctacagtct tgagatactt agaaaggctg
      601 gatggcacta atggttcacc cttagctgct g
//
```

```
>gi|22471250|gb|BU035730.1|BU035730   QHJ7N08.yg.ab1 QH_EFGHJ sunflower
RHA280 Helianthus annuus cDNA
           clone QHJ7N08.
          Length = 647

 Score =  178 bits (451), Expect = 2e-43
 Identities = 96/178 (53%), Positives = 122/178 (68%), Gaps = 3/178 (1%)
 Frame = +1

Query: 627 GLISLFSQKYFLK---SSSSFQRKDMVSSMESDVATIGSVRADDSEALPRMDARTAENIV 683
           GL++L   K+      S+S+  RK++ S++ SDV  +   R +D+E +P+MDAR AE +V
Sbjct: 16  GLMTLAGLKFIPS*TGSTSTTARKEVDSALASDVTNVEDSRVEDAEDIPKMDARLAEGLV 195

Query: 684 SKWQKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAAETAQLGLVYDYTLLKLSVDSVTV 743
            KWQ IKS A GP+H    L  VLDG M KIW  RA E AQ G  +DYTLL +++DSVTV
Sbjct: 196 RKWQSIKSQALGPEHCHSKLS*VLDGEMHKIWLQRATEIAQRGWFWDYTLLNITIDSVTV 375

Query: 744 SADGTRALVEATLEESACLSDLVHPENNATDVRTYTTRYEVFWSKSGWKITEGSVLAS 801
           S DG  A+VEATLEESA L DL HPENN +   TYTTRYE+  +KS WKIT+G+VL S
Sbjct: 376 SLDGRLAVVEATLEESAKLIDLTHPENNDSYNLTYTTRYEMSCAKSSWKITKGAVLKS 549

LOCUS       BU035730                 647 bp    mRNA    linear   EST 23-AUG-
2002
DEFINITION  QHJ7N08.yg.ab1 QH_EFGHJ sunflower RHA280 Helianthus annuus cDNA
            clone QHJ7N08, mRNA sequence.
ACCESSION   BU035730
VERSION     BU035730.1  GI:22471250
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M.,
Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig4396, see
http://cgpdb.ucdavis.edu/
            for details.
```

```
            Plate: QHJ7  row: N  column: 08.
FEATURES             Location/Qualifiers
     source          1..647
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="RHA280"
                     /db_xref="taxon:4232"
                     /clone="QHJ7N08"
                     /lab_host="E.coli"
                     /clone_lib="QH_EFGHJ sunflower RHA280"
                     /note="Vector: pBRcDNASfiAB; The library was
constructed
                     from 11 different sources of RNA from a single
genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-
fractionated,
                     directionally cloned into a custom medium-copy vector
and
                     transformations made with four size classes to
minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at
http://cgpdb.ucdavis.edu/
                     TAG_LIB=QH_EFGHJ sunflower RHA280
                     TAG_TISSUE=germinating seeds
                     TAG_SEQ=TCTGTGCGGG"
BASE COUNT      181 a    133 c    145 g    188 t
ORIGIN
        1 cagaaagagg tggctggatt gatgactttg gctggcttga aatttatacc gtcttaaaca
       61 ggctctacta gtactactgc tcgtaaagaa gttgattcgg ctctggcttc agacgtcacc
      121 aatgtggagg attctagggt tgaggatgct gaagacattc ctaaaatgga tgcaagatta
      181 gccgaaggtc tagttcgtaa gtggcagagc ataaaatccc aagcccttgg acctgagcat
      241 tgccactcaa aattatcata ggtattagat ggtgaaatgc acaagatctg gcttcaacgg
      301 gcaaccgaaa ttgctcaacg tggttggttt tgggactaca cgcttttaaa cattaccatt
      361 gacagtgtta ccgtttcact cgatgggcgc ttagctgttg tggaagcaac ccttgaagag
      421 tctgccaagt tgattgattt gacccacccg gaaaacaatg actcctataa tttaacttac
      481 accacacgtt atgagatgtc gtgtgccaag tcatcatgga aaatcacaaa gggggctgtc
      541 ctcaaatcat aacagatgta attctttctc acctttttctg tatttatctg ttattagatt
      601 actcagcagt tgaatgatat gtttctccac catttcgatc atgagcg
//
```

115

```
>gi|22394580|gb|BQ977057.1|BQ977057   QHI23M11.yg.ab1 QH_ABCDI sunflower
RHA801 Helianthus annuus cDNA
           clone QHI23M11.
          Length = 652

 Score =  166 bits (421), Expect = 5e-40
 Identities = 85/138 (61%), Positives = 101/138 (73%)
 Frame = +1

Query: 664 RADDSEALPRMDARTAENIVSKWQKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAAETA 723
           R +D+E +P+MDAR AE +V KWQ IKS A GP+H    L EVLDG M KIW  RA E A
Sbjct: 127 RVEDAEDIPKMDARLAEGLVRKWQSIKSQALGPEHCHSKLSEVLDGEMHKIWLQRATEIA 306

Query: 724 QLGLVYDYTLLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTTRYE 783
           Q G  +DYTLL +++DSVTVS DG  A+VEATLEESA L DL HPENN +   TYTTRYE
Sbjct: 307 QRGWFWDYTLLNITIDSVTVSLDGRLAVVEATLEESAKLIDLTHPENNDSYNLTYTTRYE 486

Query: 784 VFWSKSGWKITEGSVLAS 801
           +  +KS WKIT+G+VL S
Sbjct: 487 MSCAKSSWKITKGAVLKS 540


LOCUS       BQ977057                 652 bp    mRNA    linear   EST 21-AUG-
2002
DEFINITION  QHI23M11.yg.ab1 QH_ABCDI sunflower RHA801 Helianthus annuus
cDNA
            clone QHI23M11, mRNA sequence.
ACCESSION   BQ977057
VERSION     BQ977057.1  GI:22394580
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M.,
Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
```

```
           belongs to contig QH_CA_Contig4396, see
http://cgpdb.ucdavis.edu/
           for details.
           Plate: QHI23  row: M  column: 11.
FEATURES             Location/Qualifiers
     source          1..652
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="RHA801"
                     /db_xref="taxon:4232"
                     /clone="QHI23M11"
                     /lab_host="E.coli"
                     /clone_lib="QH_ABCDI sunflower RHA801"
                     /note="Vector: pBRcDNASfiAB; The library was
constructed
                     from 11 different sources of RNA from a single
genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-
fractionated,
                     directionally cloned into a custom medium-copy vector
and
                     transformations made with four size classes to
minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at
http://cgpdb.ucdavis.edu/
                     TAG_LIB=QH_ABCDI sunflower RHA801
                     TAG_TISSUE=germinating seeds
                     TAG_SEQ=TCTGTGCGGG"
BASE COUNT      178 a    135 c    148 g    191 t
ORIGIN
        1 tgtggtggtt ggattgatga ctttggctgg cttgaaattt acaccgtcca aaagaggctc
       61 tactagtact actgctcgta aagaagttga ttcggctctg gcttcagacg tcaccaatag
      121 gattctaggg ttgaggatgc tgaagacatt cctaaaatgg atgcaagatt agccgagggt
      181 ctagttcgta agtggcagag cataaaatcc caagcccttg gacctgagca ttgccactca
      241 aaattatcag aggtattaga tggtgaaatg cacaagatct ggcttcaacg ggcaaccgaa
      301 attgctcaac gtggttggtt ttgggactac acgcttttaa acattaccat tgacagtgtt
      361 accgtctcac tcgatgggcg cttagctgtt gtggaagcaa cccttgaaga gtctgccaag
      421 ttgattgatt tgacccaccc ggaaaacaat gactcctata atttaactta caccacacgt
      481 tatgagatgt cgtgtgccaa gtcttcatgg aaaatcacaa agggggctgt cctcaaatca
      541 taacagatgt aattctttct caccttttct gtatttaact gttattagat tactcagcag
      601 ttgaatgata tgtttctcca ccatatcgat catgagtgta tttggtgctg cc
//
```

```
>gi|24100065|gb|BU889000.1|BU889000   P015D07 Populus petioles cDNA library
Populus tremula cDNA 5 prime.
          Length = 460

 Score =  152 bits (384), Expect = 1e-35
 Identities = 87/149 (58%), Positives = 104/149 (69%), Gaps = 2/149 (1%)
 Frame = +1

Query: 613 KEASVKILAAGVAIGLISLFSQKYFLKSSSSFQR-KDMVSSMESDVATIGS-VRADDSEA 670
           K    + + AGVAIGL++L   K F   + SF R K++ S+M SD   + S V    SE
Sbjct: 13  KRCQYQNMCAGVAIGLLTLAGLKCFPPRTGSFIRQKEIGSAMASDTINLNSAVDEQISED 192

Query: 671 LPRMDARTAENIVSKWQKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAAETAQLGLVYD 730
           LPRMDAR AE+IV KWQ IKS AFG DH +  LPEVLD +MLKIWTDRAAE A LG VY+
Sbjct: 193 LPRMDARGAEDIVRKWQNIKSQAFGTDHCLAKLPEVLDSQMLKIWTDRAAEIAHLGWVYE 372

Query: 731 YTLLKLSVDSVTVSADGTRALVEATLEES 759
           Y LL L++DSVTVS DG  A+VEATL+ES
Sbjct: 373 YMLLDLTIDSVTVSVDGLNAVVEATLKES 459

LOCUS       BU889000                 460 bp    mRNA    linear   EST 17-OCT-
2002
DEFINITION  P015D07 Populus petioles cDNA library Populus tremula cDNA 5
prime,
            mRNA sequence.
ACCESSION   BU889000
VERSION     BU889000.1  GI:24100065
KEYWORDS    EST.
SOURCE      Populus tremula
  ORGANISM  Populus tremula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids
            ; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence
tags
            from multiple libraries
  JOURNAL   Unpublished
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1..460
                     /organism="Populus tremula"
                     /mol_type="mRNA"
                     /db_xref="taxon:113636"
                     /tissue_type="petioles"
                     /clone_lib="Populus petioles cDNA library"
```

```
BASE COUNT      138 a     82 c    117 g    123 t
ORIGIN
        1 gactgaaaaa ataaaagatg ccagtatcaa aatatgtgtg ctggtgtggc aattggactg
       61 ctgactttag ctggcctgaa gtgttttcct cctaggactg gctccttcat tcgacagaaa
      121 gaaattggtt cggcaatggc atctgacacc atcaatttga attcagcagt agatgaacaa
      181 atttccgagg acttacccag aatggatgca aggggtgcag aggatatagt tcgcaagtgg
      241 caaaacatta aatctcaggc ttttggaact gatcactgcc tggcaaaatt gccagaggtt
      301 ttggatagtc agatgttgaa aatatggaca gatcgtgcgg ccgaaattgc acatcttggt
      361 tgggtatacg agtatatgct gttggacctg actattgaca gtgtgactgt atctgtagat
      421 ggcctaaatg ctgtagtaga agcaacactc aaagagtcaa
//
```

Chlamydomonas reinhardtii ARC6-like Gene Sequence

Gene model at http://genome.jgi-psf.org/cgi-bin/dispGeneModel.v4?db=chlre1&id=140717

Genomic Sequence [46927:50859] Exons are underlined

>genie.294.6|Genomic

ATGAACTCGGCGGAGCACGTCTCTGTTGCCGTGGACTATTACCGAATGCTGCACGTTCCCCGCGTAAGCC
GCCCTGACGCCATTCGCAAGGCGTATGAGAACCTGGTGAAGCAACCCCCCGCTGCCGCGTACTCTGCGGA
CACCCTCTTCGCACGCGCGGTGCTACTCAAGGCAGCCGCGGAGTCGCTGACCGACCCGGACCTGCGCCGC
TCATATGACGCCAAGCTGGCCGCTGGTCACACAGCCCTGCGCGTCAGCCAGCAGGACCTACCCGGAGCCC
TTGTCGTGCTGCAGGAGGTGAGCCGTGCTCTGGCGACCGCTCAACCCCTTGCGACCGCTAAAACCATCAG
CACATATAGCACATATAAATTCCCATGGGTTCTGTACTACCGCCCACCCCTCTGAAGGGGGCGAGTATTC
ATTCTTCACGCATGAGCGCAGACTTTTACCCTATCAAGTCCCGCCCTCGCCCGCCTTCTCTTCCCACAGA
TCGGCGAGCACCAGTTGGTTCTGGATCTGGGTCTGCGCTGGCTAGAGGTAAACGGCGGCCAGCCCGACGC
CGGCGACGTGGCCGCTGCCGTGGCCCTGGCCTACTGTGACCGCGCTGGTGAGCGCCTCACCTCCCAGCTG
CAGCCGCCGCCGGCCTCAGCGCTGCCAGGCCCCGATGGCGCGGCGGTGCCGCACGCGCACGTGGGCGCGG
TGCTGCCCGCATGCGACGACCTGGACGCAGCGCTGAGCAAGCTCCGGCGGTACGGCATGGCGCAGCAGCT
GCAGCAGCAGATCGTGGGCGCGCTGCGGGTGAGGCTGGAGCAGGGGCTGGACCGGCAACCGGTCATAGAT
GTAGACACAGGGATGTAGGCGTCGATGCGAGGGGATGGAAGTATGGGGTCCTGTGAGTGTGAGCCGATGG
AAGGTATAGATGCTGGGAGCTGGCGCACCCGACCCATGTCATCCAAGGACTTGGCTGATGCATCGCTCAC
CCCCCGCCTCCAACCCGAATGCCCTCAGGACCTGGCGCCAGAGTACGCGTGCGAGCTGGCCGCCCTGCCG
CTGGGCGCCGAGACCGCCGCCCGGCGCGCCAAGGGCGTGGCGCTCATGCGCGGTGTGCTGCGCGCCGCCG
CCACCGTGGCCGCCGCCACAGCCAAGTAGGTGACAAGCACGCAGGAAATCGTGTGCTATATTGCATTGCG
GTACCTTGCCTTGCATCGCGGAGGCAGTGCTCGAGAATGCGTTTCGTGCGCGTGATCCGTTTGCTCGTCG
TGCCTTATCCGCCACCCCAGGCCCGAGGCTGCTGCTGACGACAGCGACGACGACGAGGTGGACCCGCGCA
GTGTGCTGGCGGCCGCCCGCCGCATGCTGACCCGCAGCCGCGACGTGCTCACCTGCAGCGAGCAGGTACA
GCGCTGCAACCGGGCAGTTATAGATGGATGCAAGTGCGTGGACGCCGAACGTACAGTTTTTGCTGTGTTC
CCCGCGTGCACCTTAGCCGCTCCTCCTGCAACCCTCACTTGCGACCTCAATGCGTGCACCTTAGCCGCTC
CTCCTGCAACCCTCAGTTGCGACCTCACGACACACCGTCTGGCTTACCCCTGCCCCCACCCCAGGTGGCC
CTGCTGCCGGACGCGCTGCGCGGCAGCGGTGTGTCGCCCACCCCGGACGCGCTGTACGACGGCGCCCTGG
CGCACCTGGTGGACGGCTTCCGCAACGGCTGGCCGCACTCCGTGCACCAGGTGGGGGAGCGCGGTGCCTG
GATGTCTGGATGGTCACTGGCCGCAAGGCTGTGCGCACCATCGGGTAGAGTGTAACCAAATGATGTGCGC
GCAATGAAGGGTGAGCAGATTCCAGCCTCCCTCTGTCGGCTGGCGTCCAACTGTGCCAACTGCGCACACA
CCTGCGCACGCCCCAGGCCGACCAGCTGCTGGCCAAGCTGGAGGCGCAGCAGGCCCGCGCAGCCGCCATG
CGCCGCGAGCAGTCCGAGCTGGCCGCCGCCGCCGCAGCCCGCCGTGCCATGTACAGCGGTCCCGCCGCCG
CCCACGGTCCCACCCTGTACACCAACTACAACAACCCTGCCGGCAGCGGCAATggcgcgccgccgccgcc
gccccgccccATGCCCATGGTGCCCAGGGGCGACGGCCAGCACGCCATGGCGGCGTCTGTGGCGGCGCAT
GTGCACTCCACGGCGATGGCGGAGCAcgcggcgcgcagcgcggctggcggcgccgccggcgccTCCGATG
GCGGCGCGCACGCCAACGGCGTGGCTCTAGAGCGGGCCGTGTGCGCCGTCCTGCTGGGTGACTACACCGC
GGCGGTGGAGCGGCTGGGGCTAGACACGAACGCGGCGGTGGAGCAGGAGCAGCTGCGCGAGTTCGTCCTG
GTGCGCCGGGGAGGGCCTACTGCAAAACGTGTTGCTCAGGGTCTTGAGATACCGAACACAATGTTTTCGT
ATACATCTCCCGTCGAGAGAGCTATGCCTCCACCGTCGGCCCGGCTCCACTGCACCCGATGCGGTTGCAG
GCCCACTCGCCCAACGGCCGCGGCGACCTGCGCCCGGGCCTGAGGGCGCTGGCCACCCGCTGGCTGGAGG
GCGTGGCGCTGGCGTCCTTCCGCGACACTGCCGGCAGCCCCGTGCCGCCGCTGGAGGCCAGCTGGTTCGC
GGACCTGCGTGTCGCCTTCTATCTGCAGGTGAGGGGCGGCAGAAGAGAGGGGGGAAAGGGAGGCGAGAAG
GCGCTTCCGCCGCTGGCGCAACGGGCCATCCTGGTGGAGCACGGCGCTACATCGCATCTGGTCCACCGTC
TCTGGATGTATAATTCGTGCACTCTTAACCGGCCGCGCAGGTATGGCGGCTGTGCCGCGTGGAGCAGGTG
CTGGCCGCCGCCCACTTCCTGGCCAACCTGCTGCCCAACATGCTCAAGgccatcgccggcactgccgtca
aggtcgcagccaacaccgccgtggcagcctcccgcgcgcagcgcctcagcgccaccgtcgcggccagcac
cgccaccgcctcgtcatcttcctctgccgcccgcggcgctcgtgccggtgccctgagcgctgccaccgcc
gccgcacacgccgcgcgccgccAGCAGGCGAACGCGGTCGGTGCCAGCATCGTCGGTGCTGACGTGCTGC
CCCCCACAGCAGTGgccgcggctgccgcggctggcacagcggccgccgccgcagtcaccggccccgccct
cggccgtggcgctgcagcttccgcctcttccttttgaggagggcgccgctgaggccgctgacctgcgtcgt
cgctttgtcgccaccagccgcggcgccagcgcggccgtcgGTGCGCCCACAGCACCAGCCGCTATGACTG
GGCCCCAGCACGGCGCCGCCTCTGCTGCGCAGTCGCACCGGGAGGAGGATGAGGATTCGCACGGCGGCCA

GGAGGGGGGCGTGCCGCGGCGCATGAGCGAGGCGGACCTGCGTGCGCACCTGGCGGGCCTGGAGAAGGCC
ATGTGGGACTCGGAGCTGCCGCCGCCGCCGCCATCCCGCGCGCAGAAGGCGCTCACCTACGCCGCAGGAC
TGGTGAGTTGCTGCGCAGCCTGACGGCCATAGTTGCCGTAGTGCCATAGTGACCGAGCACCGTGATGTTT
AGGACATGGGCGGAGAAGTGTTAGGACATGAATTGCATCAACGCTGCAAATCTGGTGTATGGTACGCGCG
TTCCCTGTCACCAACAAGGCTGTTGACCAAGCTGCTGCTGCCCTTGCACTCTTTCAACGCCCGTCTGCAG
CTGGCCGTGGTGGTGGCCTTCCTGGTGTCCAGCTTCTTccgccgcaacgacggcgccgcctccgccctgg
cacccgccgccgtcaccaccgcctccgtggccgTTAGCGCGCAGCCCGCCAAGCCGGGCAAGGCCACCCG
CTCCGCGCACTGA

Transcript Sequence [46927:50859] (without introns)

>genie.294.6|Transcript

ATGAACTCGGCGGAGCACGTCTCTGTTGCCGTGGACTATTACCGAATGCTGCACGTTCCCCGCGTAAGCC
GCCCTGACGCCATTCGCAAGGCGTATGAGAACCTGGTGAAGCAACCCCCCGCTGCCGCGTACTCTGCGGA
CACCCTCTTCGCACGCGCGGTGCTACTCAAGGCAGCCGCGGAGTCGCTGACCGACCCGGACCTGCGCCGC
TCATATGACGCCAAGCTGGCCGCTGGTCACACAGCCCTGCGCGTCAGCCAGCAGGACCTACCCGGAGCCC
TTGTCGTGCTGCAGGAGATCGGCGAGCACCAGTTGGTTCTGGATCTGGGTCTGCGCTGGCTAGAGGTAAA
CGGCGGCCAGCCCGACGCCGGCGACGTGGCCGCTGCCGTGGCCCTGGCCTACTGTGACCGCGCTGGTGAG
CGCCTCACCTCCCAGCTGCAGCCGCCGCCGGCCTCAGCGCTGCCAGGCCCCGATGGCGCGGCGGTGCCGC
ACGCGCACGTGGGCGCGGTGCTGCCCGCATGCGACGACCTGGACGCAGCGCTGAGCAAGCTCCGGCGGTA
CGGCATGGCGCAGCAGCTGCAGCAGCAGATCGTGGGCGCGCTGCGGGACCTGGCGCCAGAGTACGCGTGC
GAGCTGGCCGCCCTGCCGCTGGGCGCCGAGACCGCCGCCCGGCGCGCCAAGGGCGTGGCGCTCATGCGCG
GTGTGCTGCGCGCCGCCGCCACCGTGGCCGCCGCCACAGCCAAGCCCGAGGCTGCTGCTGACGACAGCGA
CGACGACGAGGTGGACCCGCGCAGTGTGCTGGCGGCCGCCCGCCGCATGCTGACCCGCAGCCGCGACGTG
CTCACCTGCAGCGAGCAGGTGGCCCTGCTGCCGGACGCGCTGCGCGGCAGCGGTGTGTCGCCCACCCCGG
ACGCGCTGTACGACGGCGCCCTGGCGCACCTGGTGGACGGCTTCCGCAACGGCTGGCCGCACTCCGTGCA
CCAGGCCGACCAGCTGCTGGCCAAGCTGGAGGCGCAGCAGGCCCGCGCAGCCGCCATGCGCCGCGAGCAG
TCCGAGCTGGCCGCCGCCGCCGCAGCCCGCCGTGCCATGTACAGCGGTCCCGCCGCCGCCCACGGTCCCA
CCCTGTACACCAACTACAACAACCCTGCCGGCAGCGGCAATggcgcgccgccgccgccgccccgccccAT
GCCCATGGTGCCCAGGGGCGACGGCCAGCACGCCATGGCGGCGTCTGTGGCGGCGCATGTGCACTCCACG
GCGATGGCGGAGCAcgcggcgcgcagcgcggctggcggcgccgccggcgccTCCGATGGCGGCGCGCACG
CCAACGGCGTGGCTCTAGAGCGGGCCGTGTGCGCCGTCCTGCTGGGTGACTACACCGCGGCGGTGGAGCG
GCTGGGGCTAGACACGAACGCGGCGGTGGAGCAGGAGCAGCTGCGCGAGTTCGTCCTGGCCCACTCGCCC
AACGGCCGCGGCGACCTGCGCCCGGGCCTGAGGGCGCTGGCCACCCGCTGGCTGGAGGGCGTGGCGCTGG
CGTCCTTCCGCGACACTGCCGGCAGCCCCGTGCCGCCGCTGGAGGCCAGCTGGTTCGCGGACCTGCGTGT
CGCCTTCTATCTGCAGGTATGGCGGCTGTGCCGCGTGGAGCAGGTGCTGGCCGCCGCCCACTTCCTGGCC
AACCTGCTGCCCAACATGCTCAAGgccatcgccggcactgccgtcaaggtcgcagccaacaccgccgtgg
cagcctcccgcgcgcagcgcctcagcgccaccgtcgcggccagcaccgccaccgcctcgtcatcttcctc
tgccgcccgcggcgctcgtgccggtgccctgagcgctgccaccgccgccgcacacgccgcgcgccgccAG
CAGGCGAACGCGGTCGGTGCCAGCATCGTCGGTGCTGACGTGCTGCCCCCCACAGCAGTGgccgcggctg
ccgcggctggcacagcggccgccgccgcagtcaccggccccgccctcggccgtggcgctgcagcttccgc
ctcttcctttgaggagggcgccgctgaggccgctgacctgcgtcgtcgctttgtcgccaccagccgcggc
gccagcgcggccgtcgGTGCGCCCACAGCACCAGCCGCTATGACTGGGCCCCAGCACGGCGCCGCCTCTG
CTGCGCAGTCGCACCGGGAGGAGGATGAGGATTCGCACGGCGGCCAGGAGGGGGGCGTGCCGCGGCGCAT
GAGCGAGGCGGACCTGCGTGCGCACCTGGCGGGCCTGGAGAAGGCCATGTGGGACTCGGAGCTGCCGCCG
CCGCCGCCATCCCGCGCGCAGAAGGCGCTCACCTACGCCGCAGGACTGCTGGCCGTGGTGGTGGCCTTCC
TGGTGTCCAGCTTCTTccgccgcaacgacggcgccgcctccgccctggcacccgccgccgtcaccaccgc
ctccgtggccgTTAGCGCGCAGCCCGCCAAGCCGGGCAAGGCCACCCGCTCCGCGCACTGA

Protein Sequence

>genie.294.6

MNSAEHVSVAVDYYRMLHVPRVSRPDAIRKAYENLVKQPPAAAYSADTLFARAVLLKAAAESLTDPDLRR
SYDAKLAAGHTALRVSQQDLPGALVVLQEIGEHQLVLDLGLRWLEVNGGQPDAGDVAAAVALAYCDRAGE
RLTSQLQPPPASALPGPDGAAVPHAHVGAVLPACDDLDAALSKLRRYGMAQQLQQQIVGALRDLAPEYAC
ELAALPLGAETAARRAKGVALMRGVLRAAATVAAATAKPEAAADDSDDDEVDPRSVLAAARRMLTRSRDV
LTCSEQVALLPDALRGSGVSPTPDALYDGALAHLVDGFRNGWPHSVHQADQLLAKLEAQQARAAAMRREQ

SELAAAAAARRAMYSGPAAAHGPTLYTNYNNPAGSGNGAPPPPPRPMPMVPRGDGQHAMAASVAAHVHST
AMAEHAARSAAGGAAGASDGGAHANGVALERAVCAVLLGDYTAAVERLGLDTNAAVEQEQLREFVLAHSP
NGRGDLRPGLRALATRWLEGVALASFRDTAGSPVPPLEASWFADLRVAFYLQVWRLCRVEQVLAAAHFLA
NLLPNMLKAIAGTAVKVAANTAVAASRAQRLSATVAASTATASSSSSAARGARAGALSAATAAAHAARRQ
QANAVGASIVGADVLPPTAVAAAAAAGTAAAAAVTGPALGRGAAASASSFEEGAAEAADLRRRFVATSRG
ASAAVGAPTAPAAMTGPQHGAASAAQSHREEDEDSHGGQEGGVPRRMSEADLRAHLAGLEKAMWDSELPP
PPPSRAQKALTYAAGLLAVVVAFLVSSFFRRNDGAASALAPAAVTTASVAVSAQPAKPGKATRSAH*

Thermosynechococcus elongatus BP-1 tlr0758

Location:
Init: 782410 Term: 784431 Length(aa):673
Direction: direct
Gene Products: cell division protein Ftn2 homolog

DNA sequence:
>*Thermo* (Chr) 782410-784431
GTGCGCATTCCCCTCGATTATTACCAAGTGTTGGGTGTGCCTATTCAGGCAACGCCGGAG
CAAATTGAGCAGGCCTTTCGGGACCGGCTGTTGCAGCTCCCTACCCATCAGCACTCCCCC
ACCACAGTTGCCACCCGTCGCGAACTCATTGAGCAGGCCTATGCAGTTTTGCGAGAACCG
GAGCAGCGCGATGCCTACGATCGCCACTGCCGTACCGTTGATCCCGATGATTTGATTGCC
CAGTTGGATCCCGATGCCACCACTCCCCACATTGAAATTAGTGATGAGCAATTGTCGGGG
GCACTCCTACTGCTGTATGAACTAGGAAATTATGCCCAAGTTGTCAACCTGGGAGACGCC
TTTCTTAAAAAGGATGTTTTTGAGCGCAATCGCCCCTACACTTCCCCTGCCGCCGTTGCC
GACATTACCCTCACTGTGGCTTTGGCCTATCTGGAATTGGGACGGGAGGAATGGCAGCGG
CAGTCCTATGAATCAGCCGCCTCTCAGCTAGAAGCCGGTCTCCAGGTACTTCAGCGGGTA
AATTTGTTTCCCGAGCTCCAGGAGCAGTTTCAGACGGAACTGAATCGGCTGCGTCCCTAC
CGCATTCTGGAATTACTGGCACTGCCTTTGTCCGATAGTGCGAATCGGCAGCGGGGTATT
TTATTGCTGCGGCAAATGCTGAGTGAGCGCGGGGGCATTGAGGGGCGCGGTGACGATCGC
TCAGGACTAACAGTTGAGGATTTTCTGAAATTTATTTTGCAACTGCGCAGCCATCTTACC
GTGGCAGAACAACAGGAACTCTTTGAACGGGAATCGCGGCGTCCCTCAGCGGTGGCCACC
TACCTTGCGGTACATGCCTTGGTAGCACGGGGAGTGCATGAACTGCAGCCGAGCTATATT
TGTCGGGCCAAGGATTTATTGCAGCAGCTGCTCCCCCATCAAGACGTCTATCTTGAACTT
GCCAGTTGCTTGCTGCTTTTGGGACAGCCCACCGAGGCCTTGGCAGCTCTTGACCACAGC
CAAGATCAACCGACTCTGGACTTTATCCGCCGTCATGCCGGTGAGGCTGGCGATCGACTG
CCGGGGCTTTATTACTACACCACACAATGGCTCACGGAGGAAATTTATCCTGCATTTCGG
GACTTGGGGGAAACACCCGTGGCCTTGGAGGCTTACTTTGCTGATGCCAATGTCCAAACC
TATCTAGAGGCTCTCAGTGAGGACTCCATTGCCCCTGAACCCCCTGCGACCACTGCCTCT
GCGCTCCCTGAAGTGATCAGACCAACGGTGGCCGTGCCCCCTCCCCTCTCCTTCACAGCG
GAAACGTTACCGTTGCAGGATCAGAGTCGGCTGGGTCAGGGCCTTTCGGCATCGGCTTTT
ACCCCTTCTGCAACTGCAACGGGGACATCGATGCCCCAACCATCGCCTCGCAAACGGCGC
AGCCCTCGAAACCGTTGCGCCCAAAAACGTCAGACTTGGTTTTGGATGGGTGCAGGAGTG
GTTCTTGTGGGTTTAGGGGCGTTGGCAAAAGTCTATTGGCCCGCCAAAACCGCTGAAGCC
CCCCCGCCGCCGGTGACACCGGCACCAACTCCTGTGGCAACGCCGACCCCAACGCCACAA
CCGACGACCTTAGCCATCACTTTAACACCAGAGATGGCGCGCGATCGCCTCCACACTTGG
CAGCAAATTAAAGCCCAAGCCCTTGGGCGACCATTTGAGGTGGACAAACTAACAACGATT
TTGGCGGAGCCAGAACTCAGCCGCTGGCGATCGCGGGCACAGGGCTTAAAGTCCGAGGGC
AGCTATTGGGTTTATACCCTAAAGAACTTAGAAGTGAAGGAAGTCCGCCTCCAAAGGAGC
GATCGTGTGGAGGTGTTGGCAGAAGTCAACGAGGATGCCCGTTTCTATGAACAGGGAACC
CTGCGCACTGATATTTCCTATAGCGATCCCTACCGGGTCATTTATACCTTTATCCGTCGC
GGCAATCAATGGTTGATTCAAGGCATGCAGGTGGTTAGTTAA

Protein sequence:
>tlr0758 {782410 - 784431 direct} cell division protein Ftn2 homolog
MRIPLDYYQVLGVPIQATPEQIEQAFRDRLLQLPTHQHSPTTVATRRELIEQAYAVLREPEQRDAYDRHCRTVDP
DDLIAQLDPDATTPHIEISDEQLSGALLLLYELGNYAQVVNLGDAFLKKDVFERNRPYTSPAAVADITLTVALAY
LELGREEWQRQSYESAASQLEAGLQVLQRVNLFPELQEQFQTELNRLRPYRILELLALPLSDSANRQRGILLLRQ
MLSERGGIEGRGDDRSGLTVEDFLKFILQLRSHLTVAEQQELFERESRRPSAVATYLAVHALVARGVHELQPSYI
CRAKDLLQQLLPHQDVYLELASCLLLLGQPTEALAALDHSQDQPTLDFIRRHAGEAGDRLPGLYYYTTQWLTEEI
YPAFRDLGETPVALEAYFADANVQTYLEALSEDSIAPEPPATTASALPEVIRPTVAVPPPLSFTAETLPLQDQSR
LGQGLSASAFTPSATATGTSMPQPSPRKRRSPRNRCAQKRQTWFWMGAGVVLVGLGALAKVYWPAKTAEAPPPPV
TPAPTPVATPTPTPQPTTLAITLTPEMARDRLHTWQQIKAQALGRPFEVDKLTTILAEPELSRWRSRAQGLKSEG
SYWVYTLKNLEVKEVRLQRSDRVEVLAEVNEDARFYEQGTLRTDISYSDPYRVIYTFIRRGNQWLIQGMQVVS

123

Trichodesmium erythraeum
Contig97 Gene 8639
Strand = r
Start Location: 40312
Stop Location = 37943
Stop Codon = TAA
MRNA
GTGCGGATTCCATTAGATTATTATCGAATTTTAGGTTTACCAATTCAGGCTACTGCTGAACAGTTGCGGCAGGCA
CATCA
AGACCGCACTCAGCAGTTTCCTAGAAGGGAGTATTCTGAAGCCACAATAGTTGCTCGTAAACAGCTTATAGATGA
GGCTT
ATGCTGTTCTTTGCGATCCTGAACAACGTCAAACCTATGATGGTAACTTTTTAGCTAAAACCTACGAGCCAATAG
TAGAA
GAACTCAATCCAAGTTCTCAGATAAATTTTGATCAAGCACAAGAAAAAGAAACCACACTTAAGGAGACTAGAGAA
GTTCT
TCCGGAAATAGCTTCTAAACAGTTAAAAAAAAGGACAAGTTATCAAAACAGAGAGACTAAAGCTGCCTCTGATTT
TCATT
CTAATACCCCTAGTATAGAAATAGAATATCCACAATTTGTGGGAGCCATCCTAATTTTACATGAGCTAGGAGAAT
ATGAG
CTAGTATTAAAAATAACTCACCCTTATCTTCTTAACAATAGTATAACTATTAAAGATGGACGTTTTGGAGACCCA
GCATT
AGTTTTGCCAGATGTTGTCCTTACAGTTGCTCTAGCAAATTTAGAATTGGGCAGAGAGGAATGGCAACAAGGACA
ATACG
AAAGTGCAGCTACAGCTTTAGAGGCTGGCCTAGGGTTATTGCTACGAGAAAACCTATTTGTCCAAATACGAGGAG
AGATA
CAAGCTGACCTTTATAAGCTACGTCCTTATAGAATAATGGAGCTAATAGCACTACCAGAGGAAATAGCTCTAGAC
CGTAG
CCGTGGACTAGAAATTCTTCAAGATATGCTCAATGAACGGGGAGGAATTGATGGTCAAGGTGAAGATAGCTCTGG
ACTTG
GGATAGAAGATTTTCTAAAGTTTGTTCAGCAGCTACGTCAATACTTAACTACAGCAGAGCAAAAGAAGTTATTTG
AGGCA
GAAGCCCTTCGCCCTTCCGCAGTTGGTGCATATCTAGCGGTTTATACTTTTTTAGCTCAAGGGTTTGCTCAAAAA
CAACC
AGCCTTTATTCGTAAAGCTAAGTTGATGTTAATGCAATTGGGTCGGAGTCAAGATGTAAATTTAGAGAAATCTGT
CTGTG
CTTTACTTTTAGGGCAAACTGAAGAAGCTAGTCGTTCATTAGAACTTAGCCATGAAAATGAACCTCTATCCTTTA
TTAAA
GAAAATTCTCAACAATCTCCAGATTTATTGCCAGGTCTATGTCTCTATGCTGAACATTGGTTGACAGAGGAGGTT
TTTCC
ACATTTCCGTGATTTGTCTGACAAGTCAGCTTCTTTGAAAGATTATTTTGCAGATCAACATGTTCAAGCTTATCT
AGAAG
CTTTACCTACAGAAGCAGAGGTAGCTAATCAATGGGTAGTCGTTCAGCCTCGTCGTAGTAATCACAATAAAAAAC
AAATG
TTCGACCCCAAGGAACTTGAGAAGTTGAATGTATCAGATTTGGAGGATAAAGATATTTCTCGGGTAGATGCTACT
GCTAC
TGGTATTGTTGCTTCTGGAAGTCAAGGAAGTTCTAATTTACTAGGGGCTAGTTCTGATGGGTTGCTTCAAGAATT
AGAAA
AATCATCATCTACTAGAGGTGGGCCAAAACAAGTAACTACTAAGAGTTCTAGTCACTATTTAGGAAAAATTAGGG
AAAAG
AGTATAAGTGGTTTACCTGAGTTTAATGAAAGTACATCTATTGAGAGTGGGGGGTTACCCCAATCTATCCAAGAG
CATAG
TTCACGTAGAACTTCTGCTAGAAGAGAACCTGTTAAGTTTGGTCGTTTAATATTAATCGCAATTGTGGGATTTTT
GTTAA
TAGGATTTATTGGGTTGTTAACAATTAAAACTATCGGCTGGTTAGTAAATGCTTTAGGATGGGAAAGAGAAAAAC
TGATG

ATACAATTGGATAGGCCTCCTATAGAAATCCCAGAACCTGATCGGGTTAACCTCGCAGCATCAGGACCGATAACA
AAAGA
AGTAGCAAGGCGAACAATTCAAAGTTGGTTAGATATCAAGGCTTCTGCTCTTGGTCCTAATCATAAAATTGAACA
ATTAC
CAAATATTTTAGTAGAACCGGCACTTTCTCGTTGGTTACCTACAGCTAATGCCCTGAAGCAAGAAAAGTCATACC
GTAGG
TATGAGCATGATTTAGAAATAAGTAATATAAAGATGAGTAATACAAATTCTAATCTCGCTCAAGTAGATGCTAAA
GTGAT
AGAAAAGGTAGAGTTTTATTCTGACAATGGTAGATTAACTAATACTAACAATGAAAACTTATTTGTTCGTTATGA
TTTAG
TTCGTAAAAGTCAAAAATGGCAAATTAGTAATTGGAAGGTATTGAGATAA

PROTEIN
Protein Length = 789

VRIPLDYYRILGLPIQATAEQLRQAHQDRTQQFPRREYSEATIVARKQLIDEAYAVLCDPEQRQTYDGNFLAKTY
EPIVE
ELNPSSQINFDQAQEKETTLKETREVLPEIASKQLKKRTSYQNRETKAASDFHSNTPSIEIEYPQFVGAILILHE
LGEYE
LVLKITHPYLLNNSITIKDGRFGDPALVLPDVVLTVALANLELGREEWQQGQYESAATALEAGLGLLLRENLFVQ
IRGEI
QADLYKLRPYRIMELIALPEEIALDRSRGLEILQDMLNERGGIDGQGEDSSGLGIEDFLKFVQQLRQYLTTAEQK
KLFEA
EALRPSAVGAYLAVYTFLAQGFAQKQPAFIRKAKLMLMQLGRSQDVNLEKSVCALLLGQTEEASRSLELSHENEP
LSFIK
ENSQQSPDLLPGLCLYAEHWLTEEVFPHFRDLSDKSASLKDYFADQHVQAYLEALPTEAEVANQWVVVQPRRSNH
NKKQM
FDPKELEKLNVSDLEDKDISRVDATATGIVASGSQGSSNLLGASSDGLLQELEKSSSTRGGPKQVTTKSSSHYLG
KIREK
SISGLPEFNESTSIESGGLPQSIQEHSSRRTSARREPVKFGRLILIAIVGFLLIGFIGLLTIKTIGWLVNALGWE
REKLM
IQLDRPPIEIPEPDRVNLAASGPITKEVARRTIQSWLDIKASALGPNHKIEQLPNILVEPALSRWLPTANALKQE
KSYRR
YEHDLEISNIKMSNTNSNLAQVDAKVIEKVEFYSDNGRLTNTNNENLFVRYDLVRKSQKWQISNWKVLR

Fig. 9

SEQ ID NO:11

```
56041 actgtaaatt ttgataaata aaaaaaaaca aaaaaaagat cgccaaatca tatttcatac
56101 tatcagattt aaacaatata atttgttcga cgatacagaa atattttacc tcacaggaag
56161 aggttgcgca gaaggagcca tggatgtgtt tgttcgagtc gagttgcttt gttgtaagta
56221 ggtaattgca agaaacttga gttgtctata aagctttgga atacttctct ttatatatac
56281 gtttacaaca attttttttt ttttttttt tctattttta caacaaattg ttttttatta
56341 taataataaa cttaaacgaa aataaataat atctctttgt tctatttctt aaaaaagaaa
56401 ttagcttgta gtacttcaac gtatcttaac tctttagtct ttagtaggta tatatcatct
56461 atttatttat ttttatttt tttatattac gattatagtg tacgtacgta tttattaatc
56521 aaaaataact tggtagaagt aaaaagaaaa tgatttttt tttactcagt gatcagtttt
56581 acgtttattc aaaaataagt tgtagtttcc ttcttaatat tcaagttata tgactaaaaa
56641 ttggtcggtt aatttactat taagattaat cggaaactct agttagatca cgagataatc
56701 atcacgtgga gaaacatttg gttcttgtca cgtggagaaa acgttaagct tatttttac
56761 ttctttatta tatttttgag gaaatggttg aaagaaagag agtgtttaaa atgtgaatgc
56821 gctcgtagtt aggtggaggt taatgggtag gagggtaggt catatgtgta ttagtgatgg
56881 ataaaaatta aaaacataaa aaaaacttca agctgtaaat aatctaataa aagaacatag
56941 aaatataatc aaagaaccat ttaactaaat aaatactttc gattcaaata gcatatttct
57001 aagttccaag aatagctatc ctctatccac atgttacatt tttttttct ttttcacatc
57061 catatagttt ttaaaataat tttctagatg gtatttttta ttcgacattt ttttttcctt
57121 ttagatttac tgattataat ttatttagaa ataaatgata cgactgtcgt ttctacaaaa
57181 ctgaaatttg caaacattgg accaaaaagc gaaaccttaa tcacttgaaa cgacaacgtt
57241 ctttagtatg tttttggaca tacaaagtac acataagatg ttccctcact cttcgattgt
57301 ttcttaacct aatataatta agcaatattg aacttgagtc actcaatgct gcaccgaagg
57361 agcctttaga ttttgagcaa attcatgaga gtttagcttc tcattcatca ctctgaattt
57421 ctcttttatc ctctttatct gtccaaaaca tgacacataa cataatgtta gttctcctgc
57481 atacttccaa tggcaaatag aaaaaagaga cattgatcat agaagtcagt ttggtttacc
57541 cttctgagct cgatctctgt gctccgtttc ttttgatcaa gtgattgccg gagattcgtg
57601 atgtcgaaga tactatcgag gtcgtcttca aatgcgtttt ccaactcttc ccggagaaga
57661 gcaggtaact tatcaacgat gggcattaga agaaaacagt tgaactgcag aacaaaagaa
57721 aacacagata caaacttttt aaaagaaaag tcattttaaa agcaagaaga atctgagtaa
57781 aaactgaagt aggagcaaac ctttaactca gcagaggcga gaaagtactc tcgtatgccc
57841 tggaatatct gttggaccaa tgcgtacaca attctctcag aggaaggagc aagcttgcgg
57901 ttccaaagtg tgctatctag aagatcagcc aaccgcattt ctgttgtctg aatactggaa
57961 cctgaatcga tgtttgaggc gagatggctt agctttacat ctgatcttga cttggtgtct
58021 gttgtgccac ctaatgcatc ttggggaaga ctaaatccta tggcattacc tgatgtcgta
58081 ttatgctctg ttccaccaaa tgagtccaag aattgacgta gaccagctcg gttctacata
58141 acattgagaa acgaaaacta ctcaatcaga aacggatact tgatggtatg tacacaactc
58201 aattggattg aaacagagct atagggctgt agcaatgacc ttgttgtgaa gagaccatgt
58261 aacatagcga gttgtacttg ctaaatcctc catacatctg caaacaatat aaaatccaaa
58321 gggtgatcaa tcactaaagc tcactagaac acaggtagga ggcaccgaca tggtaagaac
58381 aggaattgga aatagaatta cttgtcacga catgattttt ctgtggactc cacaaaactg
58441 ttgaatgctg aagcaacccg cttgagaaac acctcatgcc cacttaaata ttcaccttct
58501 ttctattcaa atttagaaca tacatcaaaa aatttgctgg aaagggatca tgagtatgat
58561 accgtcaaac caaagaaaac agtacctacc tgaagaagat atacagaaat tggaagcaat
58621 ctcttgagaa tgtgtagaag cctcgcccct aactatatca acgcaaaaca aacgaaaatg
58681 agaactggaa aaaactttct gtatggaaag agaaacatgt gaataacaaa atttcagatg
58741 aaagtattcc caaacatagt ttctgtaagc agaacatgtt tactcgataa ctcttatgca
58801 caaataagtt ccagcaaatc tcaaaactga atggtagtat gatttcaata tataacgtta
58861 tatttcattt ttttttttac gtacagtaca ccttaactaa ttagtaaaat tgctttccat
58921 cctccacgaa agaaaaagaa aaaagtagct atatctatgt cacctgatga aggaaaggtt
58981 caaacgtctc acgagccttc gcaactgcta taacacaagc tgttctacaa cagcaaataa
59041 gagaaagaga ataagaggcc atagaaaaca tgacaaacgt tgcagctcag attagatact
```

```
59101 gaaaggggtc tgggatgcaa agacaataaa ttgagaagtg tgttgcatgt cagtcaatcc
59161 tatgatacct ggaatagttt gttccatcat gaatatcctc aactccacat gcatttacaa
59221 tttcctccct cgttattggg ggacatttga tagcaccaac tagaaaacga aactcagcca
59281 tggcacggtg atattgtgca cccccataga gacgcatccc tgcattctgt aaaatgaaag
59341 ataatctggt tatggtctct cataattctt gaaggtccaa cgaagtatct cttttatttg
59401 tttccaatac attattcttt ggcacatatg tttcatgcgg tcaaatttat cttccatcat
59461 attataatcc atgtacaaga acaagacaac tggatttgaa gaccatgccc agcttgctct
59521 ataaagtcca acaatattct gcttcaggga aagacttacc ggtattagct tatgtgaaaa
59581 ctggagacca tcagtaccaa caaatgctcc tccttgtgtc ctttcatctt gcagtgtctc
59641 acctgaaaaa caccatgaga aattattaac aatcaaagaa cccaacataa agagaatgct
59701 gttataaaat gtgcttctgc cagtaaccaa agtatcatga ccaatgattg attgattagc
59761 atacatcatt ccatgtgtaa tcatcgcagt ctggtgaccc agtcgaattg aacaatatgc
59821 atttaactaa actgattttg caaaagtcca atttaacaac acccagaaac aagaaaagtt
59881 tatgccaaag aagttgacta gcagagaaca gagcagtaac attaccaaat ttatctggag
59941 gggccacaac tgttcccttc aataacagcg ataactgatc aagaaaaata taaacaaaac
60001 aggtgagaaa acacagcact gatcaatact aacaaaggta cttcgtacgt caatcagaaa
60061 atatgacgca gcaattttaa agtcttaagg gcatccaaca caaaaagttt acagccattc
60121 tgaatttgta gcaagtccta gatatcattt actgtagcat aattttatat gtgtcagtaa
60181 tcaataaaca aatttgtttt tatgtgtcag tagttaataa accaaaaaaa aagagaagtt
60241 tacacaaatg aacttgttgt aattatacaa aaactattaa tccacgagtc caggcaaaaa
60301 tgaaaaggta tgggaaggtg taaatagaaa tctaaaaaaa cgaaatgctc tctacagtta
60361 ccttggttaa gaagagatca tggaaagtcc tgcctctctc tttgagtttt gcttcatcca
60421 aagagctgca ttgaaaggaa ttattcaacc tccaatgagt tatattttct ataaatcagt
60481 agctaacaat taaactgcct aaaatcaagt agacattttc agacaaaaca aattgcgacc
60541 taagttcctt gctcacggta tccagctttc tgactgtact gcggtactcc tttcctaaca
60601 gtggaatgat caatggaaca ctctctttgt acctggaaag agaagggcat caagactaca
60661 gcgaaaagta aactacaata gaaacagagg ctggaaaaat cagagttaaa acaacagtta
60721 taccttttcc agagtagttc ttccagaaac aacctcagtt tactgatgcc aatcctactc
60781 ttttcctgtt ttgtcagtaa acggcccaac ttcttctcta aagatgcaat gtcttccatt
60841 tctctaagtg acacagcctg taataaaaac cacacatagt ttagaaaaag acctgtttaa
60901 cttgtttaag gaatcagaca gcagagcaga gacctgtttg aactcgtcat tagacttata
60961 cactgaatcc tgtccatagc caactcttcc agaaggcaca gacgtgaaaa aaggagaatc
61021 gcccaataag gagctgtcaa gtgcgcttgc aggaggtgag agaaagactt ccacgtcaga
61081 tgaacatgag aattgaggga ttttagtgtc aagctttgta gaaacaacaa ttgtcctaga
61141 aagctcagga tcaacctaca tgaacgagaa acaaacttta acaaaaataa agacaaggtt
61201 agacgcaatg gagttacgtc aagcaacgta cttgcatcac tatccttcga gtggttgcaa
61261 tgctccagtc actgctatct tcgaggcata aaatgatgaa ctctttgtgt tgcatctttg
61321 ctcggactag agcttccaca gcccgtgctt gaacctaaga aaaagaacaa gtaacccact
61381 ctcaaataaa gcaaaaccaa aacatgaaat cagccacgga attggctgga agccataaga
61441 aaaaacaacc tgaagagctc ggtttttcag tcctggtgca ggagcaataa gtccaggtgt
61501 atcaatgatg gtaaggtttg gacaatactt atactggact ttcacaataa tctcctttgc
61561 agagaatggg ctacatggct cttgctccag cctcatgttc tcagcctcaa tatatgccta
61621 actccaaatc atataacaaa tttcgttaac atgagcattt cgcttctcta caataaacct
61681 aagtacttgt gtttctcaac attcgtcaaa atcttcccag aatttatacg cagaaacaag
61741 caattgaaga agcacaagta ataataataa caaaacacct gaatttgtga gagagatttg
61801 ggaagagaaa cggaaggatc atcatcagat ccgagatgac aaagcgggaa ttgacactga
61861 ggatcgtact tcatatggag agtaatcggc cgacgagtct tggttccgcc gccgacatgg
61921 ttaaattgaa accccataag agcttccaca agcgcacttt taccgtcggt ctgctgtccc
61981 accacaagaa ccgccggtgc ttcgaacggc gtctccaatt cctgcgccaa agcgtgtaac
62041 tcgttgtaag cttcgtaaag actccaccgc tcctcaatcg cagcgtcgtc ctcttccgcc
62101 atttcctcaa ccgtcaccga ttttgctgat acttccgcca tcgtctctta cgaaaatgag
62161 caagaggaag agtaagagta agagagtgtc tcttatttct tctactcttt agttttcgtc
62221 gccgttcctt tttccgccat ggaattagca gatacggcta atttcaattt ttgtcaaaag
62281 aaatattttt tgtgttttaa tctcacgcgc atccatggcg cgttgagtca acgttgtaat
62341 agttctccgc taaatttaaa taaaagagcg cgtaaggaga gagtttaagg attttttttt
62401 tttggtcggc aaatacaaag gatttgcttt gtcttgacca atagtatatg cagaaatatt
```

Fig. 9, continued 3/3

```
62461 atctcaaagg atttgtgata actatgtagt acagaattgt gattattgga tgagaaacca
62521 gaaatatttt gagcaaatga cgacttgtta atttactatt ttttcatttc ttaaaggtct
62581 ctcttgtgta actatgatta aaattgaaat agtgactttt attgttacga catggaacaa
62641 atcaacgagt tctattgtta aagagagaca ttgatgaatg taacaaaact gtggcttaga
62701 agccgaaagg agacttagtt cgggtccctc cttcaccgta ttgctcgttc cattttctca
62761 attcgttcat tgtcgtcgcg tcgtatgcca ctgacggact tacctgcaaa ttacattaca
62821 atgacgcaat ttcgataatg caaacaccag gggaaaaaac atgaatagag atgatgatga
62881 tgtttttaa gagattgatc aataccttag ctttggattg aatgaagtcg tccaaactca
62941 gtggtcgtag atcaggggac gcatttgtta ccgagtcctg ataattcgac gtttcaaaag
63001 catggagtga gtacaaaaat tatttttcgt aacaacagaa atcaactgtg tgggtttatg
63061 catgtcctta ccttgttttc ttcttgtaac aattcttgaa caggtctgta tgcagctgct
63121 atgcatagat tctgcaatgt aagaaaagaa aaggaatcag aactactgtg ttgaatcata
63181 ctcgaacttg taaatgaaac cccgaatgac caaaccttta gatcgcttcc tgaatatcct
63241 tcggtttcct ttgcaagttt atcaaactcg aaaccagttt caagattttc tggtgtcaga
63301 aatatcttca atatcttcaa ccggttttcc gcatctggta aatccacata tatcctataa
63361 acacaagcct caatacaatt atcgaaaaga tacaaatatt ccaaaggaga aattacttga
63421 aagcttaaat taccgtcttg gtagcctacg aatgacagcg tcatcaagat caaaaggtcg
63481 gttggtggca ccgagaatga gaatcctttg gctatctttt gatctgagtc catcccaagc
63541 tgccataaac tcatttctca ttcttcgtgt tgcctcgtgc tcaaaagcac caccacgagc
63601 acccaacaaa ctgtcaacct atacgacaac aaaataaatt acagttagtc cttgagtaac
63661 acattttacg catcacaaaa gtattcctca taaaaagcaa taaccgaaat tgaaaagtga
63721 tataaagcta aacaatttct cacctcatca acaaatataa tgacgggggc tagtttgctt
63781 gcaaaagaga acaaagcctt cgtgagcttc tctgcatctc caaaccactg tgccaaacaa
63841 tggacgaaat tgacttaaat cagaaccaat cagaggtaaa gttggaaaga gatttactct
63901 aagttacaat cggcattgac aataataagt cgatgaccgg ggtggaaaag tttttcttat
63961 gtcattagat attctcctta tttatatgaa gatgtttaca aagtggaata tcaacgtgac
```

Fig. 10

SEQ ID NO:12

gaaa ttagccgtat ctgctaattc catggcggaa aaaggaacgg
cgacgaaaac taaagagtag aagaaataag agacactctc ttactcttac
tcttcctctt gctcattttc gtaagagacg

ATGGCGGAAGTATCAGCAAAATCGGTGACGGTTGAGGAAATGGCGGAAGA
GGACGACGCTGCGATTGAGGAGCGGTGGAGTCTTTACGAAGCTTACAACG
AGTTACACGCTTTGGCGCAGGAATTGGAGACGCCGTTCGAAGCACCGGCG
GTTCTTGTGGTGGGACAGCAGACCGACGGTAAAAGTGCGCTTGTGGAAGC
TCTTATGGGGTTTCAATTTAACCATGTCGGCGGCGGAACCAAGACTCGTC
GGCCGATTACTCTCCATATGAAGTACGATCCTCAgTGTCAATTCCCGCTT
tGTCATCTCGGATCTGATGATGATCCTTCCGTTTCTCTTCCCAAATCTCT
CTCACAAATTCACGCATATATTGAGGCTGAgAACATGAGGCTGGAGCAAG
AGCCATGTaGCCCATTCTCTGCAAAGGAGATTATTGTGAAAGTCCAGTAT
AAGTATTGTCCAAACCTTACCATCATTGATACACCTGGACTTATTGCTCC
TGCACCAGGACTGAAAAACCGAGCTCTTCAGGTTCAAGCACGGGCTGTGG
AAGCTCTAGTCCGAGCAAAGATGCAACACAAAGAGTTCATCATTTTATGC
CTCGAAGATAGCAGTGACTGGAGCATTGCAACCACTCGAAGGATAGTGAT
GCAAGTTGATCCTGAGCTTTCTAGGACAATTGTTGTTTCTACAAAGCTTG
ACACTAAAATCCCTCAATTCTCATGTTCATCTGACGTGGAAGTCTTTCTC
TCACCTCCTGCAAGCGCACTTGACAGCTCCTTATTGGGCGATTCTCCTTT
TTTCACGTCTGTGCCTTCTGGAAGAGTTGGCTATGGACAGGATTCAGTGT
ATAAGTCTAATGACGAGTTCAAACAGGCTGTGTCACTTAGAGAAATGGAA
GACATTGCATCTTTAGAGAAGAAGTTGGGCCGTTTACTGACAAAACAGGA
AAAGAGTAGGATTGGCATCAGTAAACTGAGGTTGTTTCTGGAAGAACTAC
TCTGGAAAAGGTACAAAGAGAGTGTTCCATTGATCATTCCACTGTTAGGA
AAGGAGTACCGCAGTACAGTCAGAAAGCTGGATACCGTGAGCAAGGAACT
TAGCTCTTTGGATGAAGCAAAACTCAAAGAGAGAGGCAGGACTTTCCATG
ATCTCTTCTTAACCAAGTTATCGCTGTTATTGAAGGGAACAGTTGTGGCC
CCTCCAGATAAATTTGGTGAGACACTGCAAGATGAAAGGACACAAGGAGG
AGCATTTGTTGGTACTGATGGTCTCCAGTTTTCACATAAGCTAATACaGA
ATGCAGGGATGCGTCTCTATGGGGGTGCACAATATCACCGTGCCATGGC
TGAGTTTCGTTTTCTAGTTGGTGCTATCAAATGTCCCCCAATAACGAGGG
AGGAAATTGTAAATGCATGTGGAGTTGAGGATATTCATGATGGAACAAA
CTATTCCAGAACAGCTTGTGTTATAGCAGTTGCGAAGGcTCGTgAGACGT
TTGAACcTTTCCTTCATCAGTTAGGGGCGAGGCTTCTACACATTCTCAAG
AGATTGcTTCCAATTTCTGTATATCTTCTTCAGAAAGAAGGTGAATATTT
AAGTGGGCATGAGGTGTTTCTCAAGCGGGTTGCTTCAGCATTCAACAGTT
TTGTGGAGTCCACAGAAAAATCATGTCGTGACAAATGTATGGAGGATTTA
GCAAGTACAACTCGCTATGTTACATGGTCTCTTCACAACAAGAACCGAGC
TGGTCTACGTCAATTCTTGGAcTCATTTGGTGGAACAGAGCATAATACG
ACATCAGGTAATGCCATAgGATTTAGTCTTCCCCAAGATGCATTAGGTGG

Fig. 10, continued 2/2

CACAACAGACACCAAGTCAAGATCAGATGTAAAGCTAAGCCAT
CTCGCCTCAAACATCGATTCAGGTTCCAGTATTCAGACAACAGAAATGCG
GTTGGCTGATCTTCTAGATAGCACACTTTGGAACCGCAAGCTTGCTCCTT
CCTCTGAGAGAATTGTGTACGCATTGGTCCAACAGATATTCCAGGGCATA
CGAGAGTACTTTCTCGCCTCTGCTGAGTTAAAGTTCAACTGTTTTCTTCT
AATGCCCATCGTTGATAAGTTACCTGCTCTTCTCCGGGAAGAGTTGGAAA
ACGCATTTGAAGACGACCTCGATAGTATCTTCGACATCACGAATCTCCGG
CAATCACTTGATCAAAAGAAACGGAGCACAGAGATCGAGCTCAGAAGGgT
AAAGAGGATAAAAGAGAAATTCAGAGTGATGAATGAGAAGCTAAACTCTC
ATGAATTTGCTCAAAATCTAAAGGCTCCTTCGGTGCAGCAT[redacted]

gtgact
caagttcaatattgcttaattatattaggttaagaaacaatcgaagagtg
agggaacatcttatgtgtactttgtatgtccaaaaacatactaaagaacg
ttgtcgtttcaagtgattaaggtttcgcttttggtccaatgtttgcaaa
tttcagttttgtagaaacgacagtcgtatcatttatttctaaataaatta
taatcagtaaatct

Fig. 11

SEQ ID NO:13

MAEVSAKSVTVEEMAEEDDAAIEERWSLYEAYNELHALAQELETPFEAPAVLVVGQQTDGKSALVEALMG
FQFNHVGGGTKTRRPITLHMKYDPQCQFPLCHLGSDDDPSVSLPKSLSQIHAYIEAENMRLEQEPCSPFS
AKEIIVKVQYKYCPNLTIIDTPGLIAPAPGLKNRALQVQARAVEALVRAKMQHKEFIILCLEDSSDWSIA
TTRRIVMQVDPELSRTIVVSTKLDTKIPQFSCSSDVEVFLSPPASALDSSLLGDSPFFTSVPSGRVGYGQ
DSVYKSNDEFKQAVSLREMEDIASLEKKLGRLLTKQEKSRIGISKLRLFLEELLWKRYKESVPLIIPLLG
KEYRSTVRKLDTVSKELSSLDEAKLKERGRTFHDLFLTKLSLLLKGTVVAPPDKFGETLQDERTQGGAFV
GTDGLQFSHKLIQNAGMRLYGGAQYHRAMAEFRFLVGAIKCPPITREEIVNACGVEDIHDGTNYSRTACV
IAVAKARETFEPFLHQLGARLLHILKRLLPISVYLLQKEGEYLSGHEVFLKRVASAFNSFVESTEKSCRD
KCMEDLASTTRYVTWSLHNKNRAGLRQFLDSFGGTEHNTTSGNAIGFSLPQDALGGTTDTKSRSDVKLSH
LASNIDSGSSIQTTEMRLADLLDSTLWNRKLAPSSERIVYALVQQIFQGIREYFLASAELKFNCFLLMPI
VDKLPALLREELENAFEDDLDSIFDITNLRQSLDQKKRSTEIELRRVKRIKEKFRVMNEKLNSHEFAQNL
KAPSVQH

Fig. 12

SEQ ID NO:14

```
56041 actgtaaatt ttgataaata aaaaaaaaca aaaaaaagat cgccaaatca tatttcatac
56101 tatcagattt aaacaatata atttgttcga cgatacagaa atattttacc tcacaggaag
56161 aggttgcgca gaaggagcca tggatgtgtt tgttcgagtc gagttgcttt gttgtaagta
56221 ggtaattgca agaaacttga gttgtctata aagctttgga atacttctct ttatatatac
56281 gtttacaaca attttttttt ttttttttt tctattttta caacaaattg ttttttatta
56341 taataataaa cttaaacgaa aataaataat atctctttgt tctatttctt aaaaaagaaa
56401 ttagcttgta gtacttcaac gtatcttaac tctttagtct ttagtaggta tatatcatct
56461 atttatttat ttttatttt tttatattac gattatagtg tacgtacgta tttattaatc
56521 aaaaataact tggtagaagt aaaaagaaaa tgatttttt tttactcagt gatcagtttt
56581 acgtttattc aaaaataagt tgtagtttcc ttcttaatat tcaagttata tgactaaaaa
56641 ttggtcggtt aatttactat taagattaat cggaaactct agttagatca cgagataatc
56701 atcacgtgga gaaacatttg gttcttgtca cgtggagaaa acgttaagct tatttttac
56761 ttctttatta tatttttgag gaaatggttg aaagaaagag agtgtttaaa atgtgaatgc
56821 gctcgtagtt aggtggaggt taatgggtag gagggtaggt catatgtgta ttagtgatgg
56881 ataaaaatta aaaacataaa aaaaacttca agctgtaaat aatctaataa aagaacatag
56941 aaatataatc aaagaaccat ttaactaaat aaatactttc gattcaaata gcatatttct
57001 aagttccaag aatagctatc ctctatccac atgttacatt ttttttttct ttttcacatc
57061 catatagttt ttaaaataat tttctagatg gtatttttta ttcgacattt ttttttcctt
57121 ttagatttac tgattataat ttatttagaa ataaatgata cgactgtcgt ttctacaaaa
57181 ctgaaatttg caaacattgg accaaaaagc gaaaccttaa tcacttgaaa cgacaacgtt
57241 ctttagtatg tttttggaca tacaaagtac acataagatg ttccctcact cttcgattgt
57301 ttcttaacct aatataatta agcaatattg aacttgagtc actcaatgct gcaccgaagg
57361 agcctttaga ttttgagcaa attcatgaga gtttagcttc tcattcatca ctctgaattt
57421 ctcttttatc ctctttatct gtccaaaaca tgacacataa cataatgtta gttctcctgc
57481 atacttccaa tggcaaatag aaaaaagaga cattgatcat agaagtcagt ttggtttacc
57541 cttctgagct cgatctctgt gctccgtttc ttttgatcaa gtgattgccg gagattcgtg
57601 atgtcgaaga tactatcgag gtcgtcttca aatgcgtttt ccaactcttc ccggagaaga
57661 gcaggtaact tatcaacgat gggcattaga agaaaacagt tgaactgcag aacaaaagaa
57721 aacacagata caaacttttt aaaagaaaag tcattttaaa agcaagaaga atctgagtaa
57781 aaactgaagt aggagcaaac ctttaactca gcagaggcga gaaagtactc tcgtatgccc
57841 tggaatatct gttggaccaa tgcgtacaca attctctcag aggaaggagc aagcttgcgg
57901 ttccaaagtg tgctatctag aagatcagcc aaccgcattt ctgttgtctg aatactggaa
57961 cctgaatcga tgtttgaggc gagatggctt agctttacat ctgatcttga cttggtgtct
58021 gttgtgccac ctaatgcatc ttggggaaga ctaaatccta tggcattacc tgatgtcgta
58081 ttatgctctg ttccaccaaa tgagtccaag aattgacgta gaccagctcg gttctacata
58141 acattgagaa acgaaaacta ctcaatcaga aacggatact tgatggtatg tacacaactc
58201 aattggattg aaacagagct atagggctgt agcaatgacc ttgttgtgaa gagaccatgt
58261 aacatagcga gttgtacttg ctaaatcctc catacatctg caaacaatat aaaatccaaa
58321 gggtgatcaa tcactaaagc tcactagaac acaggtagga ggcaccgaca tggtaagaac
58381 aggaattgga aatagaatta cttgtcacga catgattttt ctgtggactc cacaaaactg
58441 ttgaatgctg aagcaacccg cttgagaaac acctcatgcc cacttaaata ttcaccttct
58501 ttctattcaa atttagaaca tacatcaaaa aatttgctgg aaagggatca tgagtatgat
58561 accgtcaaac caaagaaaac agtacctacc tgaagaagat atacagaaat tggaagcaat
58621 ctcttgagaa tgtgtagaag cctcgcccct aactatatca acgcaaaaca aacgaaaatg
58681 agaactggaa aaaactttct gtatggaaag agaaacatgt gaataacaaa atttcagatg
58741 aaagtattcc caaacatagt ttctgtaagc agaacatgtt tactcgataa ctcttatgca
58801 caaataagtt ccagcaaatc tcaaaactga atggtagtat gatttcaata tataacgtta
58861 tatttcattt ttttttttac gtacagtaca ccttaactaa ttagtaaaat tgctttccat
58921 cctccacgaa agaaaaagaa aaaagtagct atatctatgt cacctgatga aggaaaggtt
58981 caaacgtctc acgagccttc gcaactgcta taacacaagc tgttctacaa cagcaaataa
59041 gagaaagaga ataagaggcc atagaaaaca tgacaaacgt tgcagctcag attagatact
```

132

Fig. 12, continued 2/3

```
59101 gaaaggggtc tgggatgcaa agacaataaa ttgagaagtg tgttgcatgt cagtcaatcc
59161 tatgatacct ggaatagttt gttccatcat gaatatcctc aactccacat gcatttacaa
59221 tttcctccct cgttattggg ggacatttga tagcaccaac tagaaaacga aactcagcca
59281 tggcacggtg atattgtgca cccccataga gacgcatccc tgcattctgt aaaatgaaag
59341 ataatctggt tatggtctct cataattctt gaaggtccaa cgaagtatct cttttatttg
59401 tttccaatac attattcttt ggcacatatg tttcatgcgg tcaaatttat cttccatcat
59461 attataatcc atgtacaaga acaagacaac tggatttgaa gaccatgccc agcttgctct
59521 ataaagtcca acaatattct gcttcaggga aagacttacc ggtattagct tatgtgaaaa
59581 ctggagacca tcagtaccaa caaatgctcc tccttgtgtc ctttcatctt gcagtgtctc
59641 acctgaaaaa caccatgaga aattattaac aatcaaagaa cccaacataa agagaatgct
59701 gttataaaat gtgcttctgc cagtaaccaa agtatcatga ccaatgattg attgattagc
59761 atacatcatt ccatgtgtaa tcatcgcagt ctggtgaccc agtcgaattg aacaatatgc
59821 atttaactaa actgattttg caaaagtcca atttaacaac acccagaaac aagaaaagtt
59881 tatgccaaag aagttgacta gcagagaaca gagcagtaac attaccaaat ttatctggag
59941 gggccacaac tgttcccttc aataacagcg ataactgatc aagaaaaata taaacaaaac
60001 aggtgagaaa acacagcact gatcaatact aacaaaggta cttcgtacgt caatcagaaa
60061 atatgacgca gcaattttaa agtcttaagg gcatccaaca caaaaagttt acagccattc
60121 tgaatttgta gcaagtccta gatatcattt actgtagcat aattttatat gtgtcagtaa
60181 tcaataaaca aatttgtttt tatgtgtcag tagttaataa accaaaaaaa aagagaagtt
60241 tacacaaatg aacttgttgt aattatacaa aaactattaa tccacgagtc caggcaaaaa
60301 tgaaaaggta tgggaaggtg taaatagaaa tctaaaaaaa cgaaatgctc tctacagtta
60361 ccttggttaa gaagagatca tggaaagtcc tgcctctctc tttgagtttt gcttcatcca
60421 aagagctgca ttgaaaggaa ttattcaacc tccaatgagt tatattttct ataaatcagt
60481 agctaacaat taaactgcct aaaatcaagt agacattttc agacaaaaca aattgcgacc
60541 taagttcctt gctcacggta tccagctttc tgactgtact gcggtactcc tttcctaaca
60601 gtggaatgat caatggaaca ctctctttgt acctggaaag agaagggcat caagactaca
60661 gcgaaaagta aactacaata gaaacagagg ctggaaaaat cagagttaaa acaacagtta
60721 taccttttcc agagtagttc ttccagaaac aacctcagtt tactgatgcc aatcctactc
60781 ttttcctgtt ttgtcagtaa acggcccaac ttcttctcta aagatgcaat gtcttccatt
60841 tctctaagtg acacagcctg taataaaaac cacacatagt ttagaaaaag acctgtttaa
60901 cttgtttaag gaatcagaca gcagagcaga gacctgtttg aactcgtcat tagacttata
60961 cactgaatcc tgtccatagc caactcttcc agaaggcaca gacgtgaaaa aaggagaatc
61021 gcccaataag gagctgtcaa gtgcgcttgc aggaggtgag agaaagactt ccacgtcaga
61081 tgaacatgag aattgaggga ttttagtgtc aagctttgta gaaacaacaa ttgtcctaga
61141 aagctcagga tcaacctaca tgaacgagaa acaaacttta acaaaaataa agacaaggtt
61201 agacgcaatg gagttacgtc aagcaacgta cttgcatcac tatccttcga gtggttgcaa
61261 tgctccagtc actgctatct tcgaggcata aaatgatgaa ctctttgtgt tgcatctttg
61321 ctcggactag agcttccaca gcccgtgctt gaacctaaga aaaagaacaa gtaacccact
61381 ctcaaataaa gcaaaaccaa aacatgaaat cagccacgga attggctgga agccataaga
61441 aaaaacaacc tgaagagctc ggtttttcag tcctggtgca ggagcaataa gtccaggtgt
61501 atcaatgatg gtaaggtttg gacaatactt atactggact ttcacaataa tctcctttgc
61561 agagaatggg ctacatggct cttgctccag cctcatgttc tcagcctcaa tatatgccta
61621 actccaaatc atataacaaa tttcgttaac atgagcattt cgcttctcta caataaacct
61681 aagtacttgt gtttctcaac attcgtcaaa atcttcccag aatttatacg cagaaacaag
61741 caattgaaga agcacaagta ataataataa caaaacacct gaatttgtga gagagatttg
61801 ggaagagaaa cggaaggatc atcatcagat ccgagatgac aaagcgggaa ttgacactga
61861 ggatcgtact tcatatggag agtaatcggc cgacgagtct tggttccgcc gccgacatgg
61921 ttaaattgaa accccataag agcttccaca agcgcacttt taccgtcggt ctgctgtccc
61981 accacaagaa ccgccggtgc ttcgaacggc gtctccaatt cctgcgccaa agcgtgtaac
62041 tcgttgtaag cttcgtaaag actccaccgc tcctcaatcg cagcgtcgtc ctcttccgcc
62101 atttcctcaa ccgtcaccga ttttgctgat acttccgcca tcgtctctta cgaaaatgag
62161 caagaggaag agtaagagta agagagtgtc tcttatttct tctactcttt agttttcgtc
62221 gccgttcctt tttccgccat ggaattagca gatacggcta atttcaattt ttgtcaaaag
62281 aaatattttt tgtgttttaa tctcacgcgc atccatggcg cgttgagtca acgttgtaat
62341 agttctccgc taaatttaaa taaaagagcg cgtaaggaga gagtttaagg attttttttt
62401 tttggtcggc aaatacaaag gatttgcttt gtcttgacca atagtatatg cagaaatatt
```

Fig. 12, continued 3/3

```
62461 atctcaaagg atttgtgata actatgtagt acagaattgt gattattgga tgagaaacca
62521 gaaatatttt gagcaaatga cgacttgtta atttactatt ttttcatttc ttaaaggtct
62581 ctcttgtgta actatgatta aaattgaaat agtgactttt attgttacga catggaacaa
62641 atcaacgagt tctattgtta aagagagaca ttgatgaatg taacaaaact gtggcttaga
62701 agccgaaagg agacttagtt cgggtccctc cttcaccgta ttgctcgttc cattttctca
62761 attcgttcat tgtcgtcgcg tcgtatgcca ctgacggact tacctgcaaa ttacattaca
62821 atgacgcaat ttcgataatg caaacaccag gggaaaaaac atgaatagag atgatgatga
62881 tgtttttaa gagattgatc aataccttag ctttggattg aatgaagtcg tccaaactca
```

Fig. 13

SEQ ID NO:15

```
   1 atggcggaag tatcagcaaa atcggtgacg gttgaggaaa tggcggaaga ggacgacgct
  61 gcgattgagg agcggtggag tctttacgaa gcttacaacg agttacacgc tttggcgcag
 121 gaattggaga cgccgttcga agcaccggcg gttcttgtgg tgggacagca gaccgacggt
 181 aaaagtgcgc ttgtggaagc tcttatgggg tttcaattta accatgtcgg cggcggaacc
 241 aagactcgtc ggccgattac tctccatatg aagtacgatc ctcagtgtca attcccgctt
 301 tgtcatctcg gatctgatga tgatccttcc gtttctcttc ccaaatctct ctcacaaatt
 361 caggcatata ttgaggctga gaacatgagg ctggagcaag agccatgtag cccattctct
 421 gcaaaggaga ttattgtgaa agtccagtat aagtattgtc caaaccttac catcattgat
 481 acacctggac ttattgctcc tgcaccagga ctgaaaaacc gagctcttca ggttcaagca
 541 cgggctgtgg aagctctagt ccgagcaaag atgcaacaca aagagttcat cattttatgc
 601 ctcgaagata gcagtgactg gagcattgca accactcgaa ggatagtgat gcaagttgat
 661 cctgagcttt ctaggacaat tgttgtttct acaaagcttg acactaaaat ccctcaattc
 721 tcatgttcat ctgacgtgga agtctttctc tcacctcctg caagcgcact tgacagctcc
 781 ttattgggcg attctccttt tttcacgtct gtgccttctg gaagagttgg ctatggacag
 841 gattcagtgt ataagtctaa tgacgagttc aaacaggctg tgtcacttag agaaatggaa
 901 gacattgcat ctttagagaa gaagttgggc cgtttactga caaaacagga aaagagtagg
 961 attggcatca gtaaactgag gttgtttctg gaagaactac tctggaaaag gtacaaagag
1021 agtgttccat tgatcattcc actgttagga aaggagtacc gcagtacagt cagaaagctg
1081 gataccttat cgctgttatt gaagggaaca gttgtggccc ctccagataa atttggtgag
1141 acactgcaag atgaaaggac acaaggagga gcatttgttg gtactgatgg tctccagttt
1201 tcacataagc taataccgaa tgcagggatg cgtctctatg ggggtgcaca atatcaccgt
1261 gccatggctg agtttcgttt tctagttggt gctatcaaat gtcccccaat aacgagggag
1321 gaaattgtaa atgcatgtgg agttgaggat attcatgatg gaacaaacta ttccagaaca
1381 gcttgtgtta tagcagttgc gaaggctcgt gagacgtttg aacctttcct tcatcagaaa
1441 gtttttttcca gttctcattt tcgtttgttt tgcgttgata tagttagggg cgaggcttct
1501 acacattctc aagagattgc ttccaatttc tgtatatctt cttcaggtag gtactgtttt
1561 ctttggtttg acggtgaata tttaagtggg catgaggtgt ttctcaagcg ggttgcttca
1621 gcattcaaca gttttgtgga gtccacagaa aaatcatgtc gtgacaaatg tatggaggat
1681 ttagcaagta caactcgcta tgttacatgg tctcttcaca acaagaaccg agctggtcta
1741 cgtcaattct tggactcatt tggtggaaca gagcataata cgacatcagg taatgccata
1801 ggatttagtc ttccccaaga tgcattaggt ggcacaacag acaccaagtc aagatcagat
1861 gtaaagctaa gccatctcgc ctcaaacatc gattcaggtt ccagtattca gacaacagaa
1921 atgcggttgg ctgatcttct agatagcaca ctttggaacc gcaagcttgc tccttcctct
1981 gagagaattg tgtacgcatt ggtccaacag atattccagg gcatacgaga gtactttctc
2041 gcctctgctg agttaaagtt caactgtttt cttctaatgc ccatcgttga taagttacct
2101 gctcttctcc gggaagagtt ggaaaacgca tttgaagacg acctcgatag tatcttcgac
2161 atcacgaatc tccggcaatc acttgatcaa aagaaacgga gcacagagat cgagctcaga
2221 aggataaaga ggataaaaga gaaattcaga gtgatgaatg agaagctaaa ctctcatgaa
2281 tttgctcaaa atctaaaggc tccttcggtg cagcattga
```

Fig. 14

SEQ ID NO:16

MAEVSAKSVTVEEMAEEDDAAIEERWSLYEAYNELHALAQELETPFEAPAVLVVGQQTD
GKSALVEALMGFQFNHVGGGTKTRRPITLHMKYDPQCQFPLCHLGSDDDPSVSLPKSLS
QIQAYIEAENMRLEQEPCSPFSAKEIIVKVQYKYCPNLTIIDTPGLIAPAPGLKNRALQ
VQARAVEALVRAKMQHKEFIILCLEDSSDWSIATTRRIVMQVDPELSRTIVVSTKLDTK
IPQFSCSSDVEVFLSPPASALDSSLLGDSPFFTSVPSGRVGYGQDSVYKSNDEFKQAVS
LREMEDIASLEKKLGRLLTKQEKSRIGISKLRLFLEELLWKRYKESVPLIIPLLGKEYR
STVRKLDTLSLLLKGTVVAPPDKFGETLQDERTQGGAFVGTDGLQFSHKLIPNAGMRLY
GGAQYHRAMAEFRFLVGAIKCPPITREEIVNACGVEDIHDGTNYSRTACVIAVAKARET
FEPFLHQKVFSSSHFRLFCVDIVRGEASTHSQEIASNFCISSSGRYCFLWFDGEYLSGH
EVFLKRVASAFNSFVESTEKSCRDKCMEDLASTTRYVTWSLHNKNRAGLRQFLDSFGGT
EHNTTSGNAIGFSLPQDALGGTTDTKSRSDVKLSHLASNIDSGSSIQTTEMRLADLLDS
TLWNRKLAPSSERIVYALVQQIFQGIREYFLASAELKFNCFLLMPIVDKLPALLREELE
NAFEDDLDSIFDITNLRQSLDQKKRSTEIELRRIKRIKEKFRVMNEKLNSHEFAQNLKA
PSVQH

Fig. 15

SEQ ID NO:17

MQELYTNRTVLNRPRFAVNVRPTRLKRNQQSQSKMQSHSKDPIN

AESRSRFEAYNRLQAAAVAFGEKLPIPEIVAIGGQSDGKSSLLEALLGFRFNVREVEM

GTRRPLILQMVHDLSALEPRCRFQISRIFFVELAILITDLDEDSEEYGSPIVSATAVA

DVIRSRTEALLKKTKTAVSPKPIVMRAEYAHCPNLTIIDTPGFVLKAKKGEPETTPDE

ILSMVKSLASPPHRILLFLQQSSVEWCSSLWLDAVREIDSSFRRTIVVVSKFDNRLKE

FSDRGEVDRYLSASGYLGENTRPYFVALPKDRSTISNDEFRRQISQVDTEVIRHLREG

VKGGFDEEKFRSCIGFGSLRDFLESELQKRYKEAAPATLALLEERCSEVTDDMLRMDM

KIQATSDVAHLRKAAMLYTASISNHVGALIDGAANPAPEQWGKTTEEERGESGIGSWP

GVSVDIKPPNAVLKLYGGAAFERVIHEFRCAAYSIECPPVSREKVANILLAHAGRGGG

RGVTEASAEIARTAARSWLAPLLDTACDRLAFVLGSLFEIALERNLNQNSEYEKKTEN

MDGYVGFHAAVRNCYSRFVKNLAKQCKQLVRHHLDSVTSPYSMACYENNYHQGGAFGA

YNKFNQASPNSFCFELSDTSRDEPMKDQENIPPEKNNGQETTPGKGGESHITVPETPS

PDQPCEIVYGLVKKEIGNGPDGVGARKRMARMVGNRNIEPFRVQNGGLMFANADNGMK

SSSAYSEICSSAAQHFARIREVLVERSVTSTLNSGFLTPCRDRLVVALGLDLFAVNDD

KFMDMFVAPGAIVVLQNERQQLQKRQKILQSCLTEFKTVARSL"

Fig. 16

SEQ ID NO:18

MANSNTYLTTPTKTPSSRRNQQSQSKMQSHSKDPINAESRSRFEAYNRLQAAAVAFGEK
LPIPEIVAIGGQSDGKSSLLEALLGFRFNVREVEMGTRRPLILQMVHDLSALEPRCRFQ
DEDSEEYGSPIVSATAVADVIRSRTEALLKKTKTAVSPKPIVMRAEYAHCPNLTIIDTP
GFVLKAKKGEPETTPDEILSMVKSLASPPHRILLFLQQSSVEWCSSLWLDAVREIDSSF
RRTIVVVSKFDNRLKEFSDRGEVDRYLSASGYLGENTRPYFVALPKDRSTISNDEFRRQ
ISQVDTEVIRHLREGVKGGFDEEKFRSCIGFGSLRDFLESELQKRYKEAAPATLALLEE
RCSEVTDDMLRMDMKIQATSDVAHLRKAAMLYTASISNHVGALIDGAANPAPEQWGKTT
EEERGESGIGSWPGVSVDIKPPNAVLKLYGGAAFERVIHEFRCAAYSIECPPVSREKVA
NILLAHAGRGGGRGVTEASAEIARTAARSWLAPLLDTACDRLAFVLGSLFEIALERNLN
QNSEYEKKTENMDGYVGFHAAVRNCYSRFVKNLAKQCKQLVRHHLDSVTSPYSMACYEN
NYHQGGAFGAYNKFNQASPNSFCFELSDTSRDEPMKDQENIPPEKNNGQETTPGKGGES
HITVPETPSPDQPCEIVYGLVKKEIGNGPDGVGARKRMARMVGNRNIEPFRVQNGGLMF
ANADNGMKSSSAYSEICSSAAQHFARIREVLVERSVTSTLNSGFLTPCRDRLVVALGLD
LFAVNDDKFMDMFVAPGAIVVLQNERQQLQKRQKILQSCLTEFKTVARSL

Fig. 17

SEQ ID NO:19

```
   1 ttcatgttct tagaagttct aaattttgat catctcttat ttgaaagctc aactaaaata
  61 gctatgatat cattccctga tgctacgtac taggttttta aattcataca cacacaaatc
 121 tataattaaa acttgttaaa ttcatacaca caaaggacaa atcttcttcg tattaaaaaa
 181 gatggaggct ctggaacatc tagtggtgcc gtatcactta cttgactggt tcaagccgtt
 241 tgtctttgtt tggaagaagt aaatttaatt gtgggagagg gatttcacga atttaaatct
 301 gtttttctcc cttttcgtgg tatactttgg acctttttgga tatgaacaca tatgtgaaaa
 361 cgttaattca tgtgtttgaa aagtaattaa tcgcgccgtc cgtcttatag ctttgggatg
 421 ggccaatagg atatttaaga gataagaaaa ctaatcagaa acacagacga aggtatctca
 481 ctctctctct ttctctctcc ATGAGAACTC TAATCTCTCA CCGGCAATGT GTGACGTCAC
 541 CGTTTCTTAT CTCCGCCGCA TCTCCACCGT TTCCTGGCCG GTGCTTTAAG TTATCCTCCT
 601 TTACTCCTCC ACGTCATAGG CGTTTTTCTT CTCTCTCGAT CAGAAACATT TCGCATGAAT
 661 CCGCCGATCA GACTTCTTCT TCTAGGCCGC GAACTCTTTA TCCTGGTGGT TACAAGCGTC
 721 CCGAACTCGC CGTTCCCGGT TTACTTCTCC GGCTAGACGC CGACGAGGTT ATGAGCGGGA
 781 ATCGTGAAGA GACTCTTGAT TTGGTCGACC GTGCTTTAGC TAAATCGGTT CAAATCGTCG
 841 TGATTGATGG CGGAGCCACC GCTGGTAAGC TCTACGAGGC GGCTTGTTTG CTGAAATCAC
 901 TTGTCAAAGG CCGTGCTTAC CTCTTGATCG CTGAACGTGT TGATATCGCC TCCGCCGTTG
 961 GTGCTAGTGG TGTTGCTCTC TCCGACGAAG gtaacaactg atttcattca gttttagcat
1021 ttaatttctc atagagtgag ttttgtctct caatgctatg tacagGTCTT CCGGCGATTG
1081 TGGCGAGAAA CACATTGATG GGATCCAACC CCGACTCGGT ACTTCTTCCA CTGGTAGCTC
1141 GGATTGTGAA GGATGTTGAT TCTGCTCTAA TTGCCTCAAG CTCCGAGGGT GCTGATTTCC
1201 TTATACTTGG ATCTGGTGAA GAAGATACGC AAGTGGCGGA TTCTTTGTTG AAGAGCGTGA
1261 AAATACCGAT ATATGTGACT TGCAGAGGCA ATGAAGAAGC TAAAGAAGAA TTGCAGTTAC
1321 TGAAATCAGG TGTTTCTGGT TTTGTTATTT CGTTGAAAGA TTTGCGTTCT TCTAGGGATG
1381 TAGCTCTTCG CCAGAGTCTT GATGGAGCTT ATGTTGTAAA TAATCATGAG ACACAAAATA
1441 TGAATGAACT GCCGGAGAAA AAGAATTCTG CTGGCTTCAT AAAATTAGAG GACAAACAGA
1501 AACTAATAGT AGAAATGGAG AAATCTGTGT TGAGAGAGAC GATTGAAATC ATCCACAAGG
1561 CGGCTCCACT Ggtgattttt atttcaaaca tttggtagtt gaagtcaatt ttttgaaatg
1621 gttctaagta ggtttttgtg tggttataat atggtttcat ttacttcttc gactattttt
1681 cattaacagA TGGAGGAAGT CTCCCTTCTA ATTGATGCTG TTTCTCGGAT CGATGAGCCG
1741 TTTCTGATGG TTATAGTGgt aattctgcac tcaactccgt caaattgtga ttccaggaat
1801 ttgcattggt attagctcta tattcattcc agaaacattt tagttacaca cttttgccag
1861 cactagatag cttgagatac aatgggcatg cttctagtca cttgtccttt agtgcttctc
1921 aatatcttct ttcgtcgcct atgactatga tgtttcgctt cttcttttgt tctgtctatg
1981 cttctcttct taatttgctt atggatctgg ttgtaaggga actgcatatt tcttaactgt
2041 accatctgct tgtgtacata gtttttcgc tttcttgtga cttgtgagta tgccgttctt
2101 ggaagatgtt ttaagtggga caagttgcct ttatgattca aaatagtttt tgtatggata
2161 attaattgga atccacaatt tgctggtact agGGGGAATT TAACTCTGGA AAATCAACGG
2221 TTATCAATGC ACTTCTTGGG AAGAGATACC TGAAAGAAGG GGTAGTCCCC ACTACCAATG
2281 AAATCACGTT TCTGTGCTAC TCTGACTTGG AATCCGAAGA GCAACAACGT TGCCAAACAC
2341 ATCCAGATGG CCAATATGTA TGCTATCTTC CTGCACCAAT ACTTAAGGAT gtgagtaatt
2401 caaaattcta ccatcgcagt cctgaatttt tactaattat ttggaggaat tgatttgggt
2461 tgttctcctt tcgagcagAT AAATATTGTT GACACACCTG GGACCAATGT GATCCTTCAA
2521 AGGCAACAGC GTCTTACAGA AGAATTTGTT CCACGTGCAG ATTTGCTTGT TTTTGTTCTT
2581 TCTGCTGACC GCCCTTTAAC TGAAAGTGAG gtagaagtta ccgttttact tggcatgtta
2641 gttgttgttg tttttgctca atatgtatct gcctaagtag cttgttagat ctatttttca
2701 cgaaagtagt tagttaagtc atgtatagac catcaagacc ttgtgtaggg aagggaaagt
2761 tgtcactagg ttgaatgcat atatcaaggt tttgttgatt ataaatttaa actagactaa
2821 tttattttca aagtaatgag tgttatagct attgctggaa ccagtatgtc ctgttggtcc
2881 atattttggt aaagcttagg ccaatacatt tgagaggtga gttgttattg gtacagcaaa
2941 actgatttta cgtccatggc aaattgtatg taaatgatca tctacgaagt actaacctta
```

Fig. 17, continued 2/2

```
3001 tgaatatttg gttcttattt tgaaaatctg aaaaagtttc aaaagaagga ataagcttct
3061 caatgtcatc atacccatgt catttctatc tctacctctg gagcttcctg ctgtcttgat
3121 tttactgtag gctgatttac atctcattgc gtttgtcagG TTGCGTTTCT CCGGTACACA
3181 CAGCAGTGGA AAAAGAAATT TGTGTTTATT CTGAATAAAT CTGATATCTA TCGTGATGCT
3241 CGTGAGgttt atcagaaaca atatttatgt cttttccttg atagtctctg taattgctgg
3301 atttttcttg actaaagatt aattttactg ctgcagCTTG AGGAAGCTAT TTCATTTGTT
3361 AAAGAGAATA CACGGAAGTT GCTTAATACA GAAAATGTGA TATTGTATCC GGTGTCCGCA
3421 CGGTCTGCTC TTGAGGCGAA GCTTTCAACA GCTTCTTTGG TTGGCAGAGA TGATCTTGAG
3481 ATCGCAGATC CTGGTTCTAA TTGGAGAGTC CAGAGCTTCA ATGAACTTGA GAAATTTCTT
3541 TATAGCTTCT TGGATAGCTC AACAGCTACC GGGATGGAGA GAATAAGGCT TAAATTGGAG
3601 ACACCCATGG CGATTGCTGA GCGTCTCCTT TCTTCTGTGG AAGCTCTTGT GAGACAAGAT
3661 TGCCTAGCTG CTAGGGAAGA CTTGGCTTCA GCAGACAAGA TTATCAGTCG AACTAAAGAA
3721 TACGCGCTTA AGATGGAATA TGAGAGCATT TCTTGGAGAA GGCAGGCTCT CTCGTTGGTA
3781 TAAattctat tagatattat cttgttgaat cacgaaggag gaaattggat tgttctaact
3841 tggctttttt gtgttttgta ctctggcttt tatcgcagat tgataatgcc agattacaag
3901 ttgttgatct gataggaact accctgcgac tatcaagcct tgatcttgcg atctcgtacg
3961 tgttcaaagg ggaaaaatcg gcctcagtag cagctacatc caaagttcaa ggtgaaatac
4021 tcgctccagc actcacaaat gcgaaagtaa gtgtgatgct ttattctttg agtattggcc
4081 taactgggga catgttggtc atatatatga ggtctgagat atagtcacta ttcatgcaga
4141 aagtaaatat tgtctaacaa tgtcttgttg tgacctgatt gactttacat ttcactgttt
4201 gcaggaattg cttggaaaat atgctgaatg gctacaatca aatactgccc gtgaagggag
4261 tctgtctctg aaatcattcg aaa
```

Fig. 18

SEQ ID NO:20

```
   1 ATGAGAACTC TAATCTCTCA CCGGCAATGT GTGACGTCAC CGTTTCTTAT CTCCGCCGCA
  61 TCTCCACCGT TTCCTGGCCG GTGCTTTAAG TTATCCTCCT TTACTCCTCC ACGTCATAGG
 121 CGTTTTTCTT CTCTCTCGAT CAGAAACATT TCGCATGAAT CCGCCGATCA GACTTCTTCT
 181 TCTAGGCCGC GAACTCTTTA TCCTGGTGGT TACAAGCGTC CCGAACTCGC CGTTCCCGGT
 241 TTACTTCTCC GGCTAGACGC CGACGAGGTT ATGAGCGGGA ATCGTGAAGA GACTCTTGAT
 301 TTGGTCGACC GTGCTTTAGC TAAATCGGTT CAAATCGTCG TGATTGATGG CGGAGCCACC
 361 GCTGGTAAGC TCTACGAGGC GGCTTGTTTG CTGAAATCAC TTGTCAAAGG CCGTGCTTAC
 421 CTCTTGATCG CTGAACGTGT TGATATCGCC TCCGCCGTTG GTGCTAGTGG TGTTGCTCTC
 481 TCCGACGAAG GTCTTCCGGC GATTGTGGCG AGAAACACAT TGATGGGATC CAACCCCGAC
 541 TCGGTACTTC TTCCACTGGT AGCTCGGATT GTGAAGGATG TTGATTCTGC TCTAATTGCC
 601 TCAAGCTCCG AGGGTGCTGA TTTCCTTATA CTTGGATCTG GTGAAGAAGA TACGCAAGTG
 661 GCGGATTCTT TGTTGAAGAG CGTGAAAATA CCGATATATG TGACTTGCAG AGGCAATGAA
 721 GAAGCTAAAG AAGAATTGCA GTTACTGAAA TCAGGTGTTT CTGGTTTTGT TATTTCGTTG
 781 AAAGATTTGC GTTCTTCTAG GGATGTAGCT CTTCGCCAGA GTCTTGATGG AGCTTATGTT
 841 GTAAATAATC ATGAGACACA AAATATGAAT GAACTGCCGG AGAAAAAGAA TTCTGCTGGC
 901 TTCATAAAAT TAGAGGACAA ACAGAAACTA ATAGTAGAAA TGGAGAAATC TGTGTTGAGA
 961 GAGACGATTG AAATCATCCA CAAGGCGGCT CCACTGATGG AGGAAGTCTC CCTTCTAATT
1021 GATGCTGTTT CTCGGATCGA TGAGCCGTTT CTGATGGTTA TAGTGGGGGA ATTTAACTCT
1081 GGAAAATCAA CGGTTATCAA TGCACTTCTT GGGAAGAGAT ACCTGAAAGA AGGGGTAGTC
1141 CCCACTACCA ATGAAATCAC GTTTCTGTGC TACTCTGACT TGGAATCCGA AGAGCAACAA
1201 CGTTGCCAAA CACATCCAGA TGGCCAATAT GTATGCTATC TTCCTGCACC AATACTTAAG
1261 GATATAAATA TTGTTGACAC ACCTGGGACC AATGTGATCC TTCAAAGGCA ACAGCGTCTT
1321 ACAGAAGAAT TTGTTCCACG TGCAGATTTG CTTGTTTTTG TTCTTTCTGC TGACCGCCCT
1381 TTAACTGAAA GTGAGGTTGC GTTTCTCCGG TACACACAGC AGTGGAAAAA GAAATTTGTG
1441 TTTATTCTGA ATAAATCTGA TATCTATCGT GATGCTCGTG AGCTTGAGGA AGCTATTTCA
1501 TTTGTTAAAG AGAATACACG GAAGTTGCTT AATACAGAAA ATGTGATATT GTATCCGGTG
1561 TCCGCACGGT CTGCTCTTGA GGCGAAGCTT TCAACAGCTT CTTTGGTTGG CAGAGATGAT
1621 CTTGAGATCG CAGATCCTGG TTCTAATTGG AGAGTCCAGA GCTTCAATGA ACTTGAGAAA
1681 TTTCTTTATA GCTTCTTGGA TAGCTCAACA GCTACCGGGA TGGAGAGAAT AAGGCTTAAA
1741 TTGGAGACAC CCATGGCGAT TGCTGAGCGT CTCCTTTCTT CTGTGGAAGC TCTTGTGAGA
1801 CAAGATTGCC TAGCTGCTAG GGAAGACTTG GCTTCAGCAG ACAAGATTAT CAGTCGAACT
1861 AAAGAATACG CGCTTAAGAT GGAATATGAG AGCATTTCTT GGAGAAGGCA GGCTCTCTCG
1921 TTGGTATAA
```

Fig. 19

SEQ ID NO:21

MRTLISHRQC VTSPFLISAA SPPFPGRCFK LSSFTPPRHR RFSSLSIRNI SHESADQTSS
SRPRTLYPGG YKRPELAVPG LLLRLDADEV MSGNREETLD LVDRALAKSV QIVVIDGGAT
AGKLYEAACL LKSLVKGRAY LLIAERVDIA SAVGASGVAL SDEGLPAIVA RNTLMGSNPD
SVLLPLVARI VKDVDSALIA SSSEGADFLI LGSGEEDTQV ADSLLKSVKI PIYVTCRGNE
EAKEELQLLK SGVSGFVISL KDLRSSRDVA LRQSLDGAYV VNNHETQNMN ELPEKKNSAG
FIKLEDKQKL IVEMEKSVLR ETIEIIHKAA PLMEEVSLLI DAVSRIDEPF LMVIVGEFNS
GKSTVINALL GKRYLKEGVV PTTNEITFLC YSDLESEEQQ RCQTHPDGQY VCYLPAPILK
DINIVDTPGT NVILQRQQRL TEEFVPRADL LVFVLSADRP LTESEVAFLR YTQQWKKKFV
FILNKSDIYR DARELEEAIS FVKENTRKLL NTENVILYPV SARSALEAKL STASLVGRDD
LEIADPGSNW RVQSFNELEK FLYSFLDSST ATGMERIRLK LETPMAIAER LLSSVEALVR
QDCLAAREDL ASADKIISRT KEYALKMEYE SISWRRQALS LV

Fig. 20

SEQ ID NO:22

```
   1 actgtcacaa agaactagaa aaggcaagca aaactcaact atgtcaaaag tgtcacttag
  61 attgattctt gaatagcgag acgaagtatc tgggaaaata cggtactgaa ttaacatctc
 121 cgtcagatca taggttcgga ttgaacagat gacacaatta aacaatgatg aagatcaaga
 181 cactttaatc gactgaattc tagttagaac ttagactaaa agtatttaat acttgaagct
 241 caccacttct cgaatatctt gttccaatcg ttttgatgtg gttccggcac tcaagttctg
 301 tattgttttc aagctgactt tatcagtttt ctgaagtaag tcatatgtgt ctatgcccaa
 361 ttgcgttttt gaattgacat atgttggcca tttgttttcg aatgatttca gagacagact
 421 cccttcacgg gcagtatttg attgtagcca ttcagcatat tttccaagca attcctgcaa
 481 acagtgaaat gtaaagtcaa tcaggtcaca acaagacatt gttagacaat atttactttc
 541 tgcatgaata gtgactatat ctcagacctc atatatatga ccaacatgtc cccagttagg
 601 ccaatactca aagaataaag catcacactt actttcgcat ttgtgagtgc tggagcgagt
 661 atttcacctt gaactttgga tgtagctgct actgaggccg atttttcccc tttgaacacg
 721 tacgagatcg caagatcaag gcttgatagt cgcagggtag ttcctatcag atcaacaact
 781 tgtaatctgg cattatcaat ctgcgataaa agccagagta caaaacacaa aaaagccaag
 841 ttagaacaat ccaatttcct ccttcgtgat tcaacaagat aatatctaat agaatttata
 901 ccaacgagag agcctgcctt ctccaagaaa tgctctcata ttccatctta agcgcgtatt
 961 ctttagttcg actgataatc ttgtctgctg aagccaagtc ttccctagca gctaggcaat
1021 cttgtctcac aagagcttcc acagaagaaa ggagacgctc agcaatcgcc atgggtgtct
1081 ccaatttaag ccttattctc tccatcccgg tagctgttga gctatccaag aagctataaa
1141 gaaatttctc aagttcattg aagctctgga ctctccaatt agaaccagga tctgcgatct
1201 caagatcatc tctgccaacc aaagaagctg ttgaaagctt cgcctcaaga gcagaccgtg
1261 cggacaccgg atacaatatc acattttctg tattaagcaa cttccgtgta ttctctttaa
1321 caaatgaaat agcttcctca agctgcagca gtaaaattaa tctttagtca agaaaaatcc
1381 agcaattaca gagactatca aggaaaagac ataaatattg tttctgataa acctcacgag
1441 catcacgata gatatcagat ttattcagaa taaacacaaa tttctttttc cactgctgtg
1501 tgtaccggag aaacgcaacc tgacaaacgc aatgagatgt aaatcagcct acagtaaaat
1561 caagacagca ggaagctcca gaggtagaga tagaaatgac atgggtatga tgacattgag
1621 aagcttattc cttcttttga aacttttca gattttcaaa ataagaacca aatattcata
1681 aggttagtac ttcgtagatg atcatttaca tacaatttgc catggacgta aaatcagttt
1741 tgctgtacca ataacaactc acctctcaaa tgtattggcc taagctttac caaaatatgg
1801 accaacagga catactggtt ccagcaatag ctataacact cattactttg aaaataaatt
1861 agtctagttt aaatttataa tcaacaaaac cttgatatat gcattcaacc tagtgacaac
1921 tttcccttcc ctacacaagg tcttgatggt ctatacatga cttaactaac tactttcgtg
1981 aaaaatagat ctaacaagct acttaggcag atacatattg agcaaaaaca acaacaacta
2041 acatgccaag taaaacggta acttctacct cactttcagt taaagggcgg tcagcagaaa
2101 gaacaaaaac aagcaaatct gcacgtggaa caaattcttc tgtaagacgc tgttgccttt
2161 gaaggatcac attggtccca ggtgtgtcaa caatatttat ctgctcgaaa ggagaacaac
2221 ccaaatcaat tcctccaaat aattagtaaa aattcaggac tgcgatggta gaattttgaa
2281 ttactcacat ccttaagtat tggtgcagga agatagcata catattggcc atctggatgt
2341 gtttggcaac gttgttgctc ttcggattcc aagtcagagt agcacagaaa cgtgatttca
2401 ttggtagtgg ggactacccc ttctttcagg tatctcttcc caagaagtgc attgataacc
2461 gttgattttc cagagttaaa ttccccctag taccagcaaa ttgtggattc caattaatta
2521 tccatacaaa aactattttg aatcataaag gcaacttgtc ccacttaaaa catcttccaa
2581 gaacggcata ctcacaagtc acaagaaagc gaaaaaacta tgtacacaag cagatggtac
2641 agttaagaaa tatgcagttc ccttacaacc agatccataa gcaaattaag aagagaagca
2701 tagacagaac aaaagaagaa gcgaaacatc atagtcatag gcgacgaaag aagatattga
2761 gaagcactaa aggacaagtg actagaagca tgcccattgt atctcaagct atctagtgct
2821 ggcaaaagtg tgtaactaaa atgtttctgg aatgaatata gagctaatac caatgcaaat
2881 tcctggaatc acaatttgac ggagttgagt gcagaattac cactataacc atcagaaacg
2941 gctcatcgat ccgagaaaca gcatcaatta gaagggagac ttcctccatc tgttaatgaa
```

Fig. 20, continued 2/2

```
3001 aaatagtcga agaagtaaat gaaaccatat tataaccaca caaaaaccta cttagaacca
3061 tttcaaaaaa ttgacttcaa ctaccaaatg tttgaaataa aaatcaccag tggagccgcc
3121 ttgtggatga tttcaatcgt ctctctcaac acagatttct ccatttctac tattagtttc
3181 tgtttgtcct ctaattttat gaagccagca gaattctttt tctccggcag ttcattcata
3241 ttttgtgtct catgattatt tacaacataa gctccatcaa gactctggcg aagagctaca
3301 tccctagaag aacgcaaatc tttcaacgaa ataacaaaac cagaaacacc tgatttcagt
3361 aactgcaatt cttctttagc ttcttcattg cctctgcaag tcacatatat cggtattttc
3421 acgctcttca acaaagaatc cgccacttgc gtatcttctt caccagatcc aagtataagg
3481 aaatcagcac cctcggagct tgaggcaatt agagcagaat caacatcctt cacaatccga
3541 gctaccagtg gaagaagtac cgagtcgggg ttggatccca tcaatgtgtt tctcgccaca
3601 atcgccggaa gacctgtaca tagcattgag agacaaaact cactctatga gaaattaaat
3661 gctaaaactg aatgaaatca gttgttacct tcgtcggaga gagcaacacc actagcacca
3721 acggcggagg cgatatcaac acgttcagcg atcaagaggt aagcacggcc tttgacaagt
3781 gatttcagca aacaagccgc ctcgtagagc ttaccagcgg tggctccgcc atcaatcacg
3841 acgatttgaa ccgatttagc taaagcacgg tcgaccaaat caagagtctc ttcacgattc
3901 ccgctcataa cctcgtcggc gtctagccgg agaagtaaac cgggaacggc gagttcggga
3961 cgcttgtaac caccaggata aagagttcgc ggcctagaag aagaagtctg atcggcggat
4021 tcatgcgaaa tgtttctgat cgagagagaa gaaaaacgcc tatgacgtgg aggagtaaag
4081 gaggataact taaagcaccg gccaggaaac ggtggagatg cggcggagat aagaaacggt
4141 gacgtcacac attgccggtg agagattaga gttctcatgg agagagaaag agagagagtg
4201 agataccttc gtctgtgttt ctgattagtt ttcttatctc ttaaatatcc tattggccca
4261 tcccaaagct ataagacgga cggcgcgatt aattactttt caaacacatg aattaacgtt
4321 ttcacatatg tgttcatatc caaaaggtcc aaagtatacc acgaaaaggg agaaaaacag
4381 atttaaattc gtgaaatccc tctcccacaa ttaaatttac ttcttccaaa caaagacaaa
4441 cggcttgaac cagtcaagta agtgatacgg caccactaga tgttccagag cctccatctt
4501 ttttaatacg aagaagattt gtcctttgtg tgtatgaatt taacaagttt taattataga
4561 tttgtgtgtg tatgaattta aaaacctagt acgtagcatc agggaatgat atcatagcta
4621 ttttagttga gctttcaaat aagagatgat caaaatttag aacttctaag aacatgaacg
4681 aataaacaac tattttcttt tcaaaccaac taaggtagat ggtcactgaa agtatataca
4741 tcagataaaa gttgcttgtt attccagatg aagttggacc gagaaaaaaa aaagttactt
4801 gttattcaat atgtttggat ctttgtcttg cagattgcta tatagggttg ataatgggct
4861 tcgttgtaat gggtatacag tgtataagaa tcggccttgt gcaaccaatc ctaatatgtg
4921 tgtctcatta aggtaagtgc ttaagattag aagagtaaaa cacttgactt atcaactatg
4981 tcaactaagg gttctatatt tttattaaat aaaaaataat tgaatatttt ttagaatgat
5041 ttaataaatt taatgctatt gtttgattta aatgtataat tcaccgcgag aagaaatttt
5101 ataactcaaa ttttaaagtt ttaagttgta tttgtttatt ttgttaaatg tttaatattg
5161 tataattgta ttttgattgt tgtttctcgg atttcacccg tagtacatca tcccatatta
5221 atatcgaatc aaacccgtca attctaaaat ttcacccgtg gtagtattta attgtataat
5281 tatattttaa ttgtcattct aagatttcac tcctaattct atcgcaaatt attatcaacc
5341 caaaccagtc aattctaaaa tatcacccgt agtacaccat cccatattaa tatcgaatca
5401 agcccgtcaa ttctaggatt tcacccgtgg tagtatttaa ttgtataatt atattttaat
5461 tgtcattcta ggatttcact cctaattcta tcgcaaatta ttatcaaccc aaaccagtca
5521 attctaaaat atcacccgta gtacaccatc ccatattaat atcgattcaa actcgtcaat
5581 tctaggattt cgctcgtggt agtatttaat tgtataatta tattttaatt gtcattttaa
5641 ctcctagttc tatcgcaaat tcttatcaac ccaaacagtc aattctaaaa tttcacccgt
5701 agtataaagt ttaaatattt ataatattta aatttcttat aaaagaatca aaatgtgttt
5761 taaaaaaatt aaagttttaa gttttttttt tttaatattg ttaattttgt ttagtgttta
5821 agattatata attacattat gattgtcatt atatgttttt ctccatagca tactatccca
5881 tgttattatc cactcaaacc tgtcacacca tataaccccg tcccgtgaaa ttaaacacaa
5941 atttgtcatt ttattataaa tttcaaatat ttataaaatt agaaacttca aaaaagatta
6001 atattgaccc aaacttcatc attgaatttt gagtgttata tctaagattt ctctcgcaat
```

Fig. 21

SEQ ID NO:23

```
   1 atggaggctc tggaacatct agtgctttgg gatgggccaa taggatattt aagagataag
  61 aaaactaatc agaaacacag acgaaggtat ctcactctct ctctttctct ctccatgaga
 121 actctaatct ctcaccggca atgtgtgacg tcaccgtttc ttatctccgc cgcatctcca
 181 ccgtttcctg gccggtgctt taagttatcc tcctttactc ctccacgtca taggcgtttt
 241 tcttctctct cgatcagaaa catttcgcat gaatccgccg atcagacttc ttcttctagg
 301 ccgcgaactc tttatcctgg tggttacaag cgtcccgaac tcgccgttcc cggtttactt
 361 ctccggctag acgccgacga ggttatgagc gggaatcgtg aagagactct tgatttggtc
 421 gaccgtgctt tagctaaatc ggttcaaatc gtcgtgattg atggcggagc caccgctggt
 481 aagctctacg aggcggcttg tttgctgaaa tcacttgtca aaggccgtgc ttacctcttg
 541 atcgctgaac gtgttgatat cgcctccgcc gttggtgcta gtggtgttgc tctctccgac
 601 gaaggtcttc cggcgattgt ggcgagaaac acattgatgg gatccaaccc cgactcggta
 661 cttcttccac tggtagctcg gattgtgaag gatgttgatt ctgctctaat tgcctcaagc
 721 tccgagggtg ctgatttcct tatacttgga tctggtgaag aagatacgca agtggcggat
 781 tctttgttga agagcgtgaa aataccgata tatgtgactt gcagaggcaa tgaagaagct
 841 aaagaagaat tgcagttact gaaatcaggt gtttctggtt ttgttatttc gttgaaagat
 901 ttgcgttctt ctagggatgt agctcttcgc cagagtcttg atggagctta tgttgtaaat
 961 aatcatgaga cacaaaatat gaatgaactg ccggagaaaa agaattctgc tggcttcata
1021 aaattagagg acaaacagaa actaatagta gaaatggaga aatctgtgtt gagagagacg
1081 attgaaatca tccacaaggc ggctccactg atggaggaag tctcccttct aattgatgct
1141 gtttctcgga tcgatgagcc gtttctgatg gttatagtgg gggaatttaa ctctggaaaa
1201 tcaacggtta tcaatgcact tcttgggaag agatacctga aagaaggggt agtccccact
1261 accaatgaaa tcacgtttct gtgctactct gacttggaat ccgaagagca acaacgttgc
1321 caaacacatc cagatggcca atatataaat attgttgaca cacctgggac caatgtgatc
1381 cttcaaaggc aacagcgtct tacagaagaa tttgttccac gtgcagattt gcttgttttt
1441 gttctttctg ctgaccgccc tttaactgaa agtgaggtag aagttaccgt tttacttggc
1501 atggaaggga aagttgtcac taggttgaat gcatatatca aggttgcgtt tctccggtac
1561 acacagcagt ggaaaaagaa atttgtgttt attctgaata aatctgatat ctatcgtgat
1621 gctcgtgagc ttgaggaagc tatttcattt gttaaagaga atacacggaa gttgcttaat
1681 acagaaaatg tgatattgta tccggtgtcc gcacggtctg ctcttgaggc gaagctttca
1741 acagcttctt tggttggcag agatgatctt gagatcgcag atcctggttc taattggaga
1801 gtccagagct tcaatgaact tgagaaattt ctttatagct tcttggatag ctcaacagct
1861 accgggatgg agagaataag gcttaaattg gagacaccca tggcgattgc tgagcgtctc
1921 ctttcttctg tggaagctct tgtgagacaa gattgcctag ctgctaggga agacttggct
1981 tcagcagaca agattatcag tcgaactaaa gaatacgcgc ttaagatgga atatgagagc
2041 atttcttgga gaaggcaggc tctctcgttg attgataatg ccagattaca agttgttgat
2101 ctgataggaa ctaccctgcg actatcaagc cttgatcttg cgatctcgta cgtgttcaaa
2161 ggggaaaaat cggcctcagt agcagctaca tccaaagttc aaggtgaaat actcgctcca
2221 gcactcacaa atgcgaaaga attgcttgga aaatatgctg aatggctaca atcaaatact
2281 gcccgtgaag ggagtctgtc tctgaaatca ttcgaaaaca aatggccaac atatgtcaat
2341 tcaaaaacgc aattgggcat agacacatat gacttacttc agaaaactga taaagtcagc
2401 ttgaaaacaa tacagaactt gagtgccgga accacatcaa aacgattgga acaagatatt
2461 cgagaagtg
```

Fig. 22

SEQ ID NO:24

MEALEHLVLWDGPIGYLRDKKTNQKHRRRYLTLSLSLSMRTLISHRQCVTSPFLISAASPPFPGRCFKLSS
FTPPRHRRFSSLSIRNISHESADQTSSSRPRTLYPGGYKRPELAVPGLLLRLDADEVMSGNREETLDLVDR
ALAKSVQIVVIDGGATAGKLYEAACLLKSLVKGRAYLLIAERVDIASAVGASGVALSDEGLPAIVARNTLM
GSNPDSVLLPLVARIVKDVDSALIASSSEGADFLILGSGEEDTQVADSLLKSVKIPIYVTCRGNEEAKEEL
QLLKSGVSGFVISLKDLRSSRDVALRQSLDGAYVVNNHETQNMNELPEKKNSAGFIKLEDKQKLIVEMEKS
VLRETIEIIHKAAPLMEEVSLLIDAVSRIDEPFLMVIVGEFNSGKSTVINALLGKRYLKEGVVPTTNEITF
LCYSDLESEEQQRCQTHPDGQYINIVDTPGTNVILQRQQRLTEEFVPRADLLVFVLSADRPLTESEVEVTV
LLGMEGKVVTRLNAYIKVAFLRYTQQWKKKFVFILNKSDIYRDARELEEAISFVKENTRKLLNTENVILYP
VSARSALEAKLSTASLVGRDDLEIADPGSNWRVQSFNELEKFLYSFLDSSTATGMERIRLKLETPMAIAER
LLSSVEALVRQDCLAAREDLASADKIISRTKEYALKMEYESISWRRQALSLIDNARLQVVDLIGTTLRLSS
LDLAISYVFKGEKSASVAATSKVQGEILAPALTNAKELLGKYAEWLQSNTAREGSLSLKSFENKWPTYVNS
KTQLGIDTYDLLQKTDKVSLKTIQNLSAGTTSKRLEQDIREV

Fig. 23

SEQ ID NO:25

```
69061 acaaagacca gttaaaaacg tgtgtagtat aacttactgg taagtaaagc tataagcaag
69121 aatctgtacc ttatttttctc tctctctagt gagccctgac catccgaatt tcgcattcgc
69181 caatcgctgt gtttccgtgt gttttccccc tttttggttt tagatttgcc taaaccaatc
69241 agaacaagag aaacctggaa acaagaacca aaaaaagtgg gctttctctg catcatcatt
69301 ccacttctgg tccccaactg aaaaggacaa tccaaagcta gatcccttca aattttcctt
69361 tttgttttcg aaattttcgc aatttttaat attattttgg aagtctatgt ttctttctga
69421 tctttagcaa caaaggaagg tggaatctgt ttcacgttta cacaaaaaca tgtcaactgg
69481 agattttctc tttccctaac ttttgaccat acagtatggt ccatacttaa tattctctct
69541 ttgtttttaa taaaataaaa ggtttggtta tcaagcatat atgtcattag cttaaagcta
69601 tgactttgtt tagaaaactt aggaggacca tatggcaagc ttttatacag tgttagactt
69661 ctaacgttaa ttctaaacaa tctccagtat caagcattaa caaggtttat tctagcacct
69721 ctggattttt aaaacttctc gaaccaatcc ttaactaaaa aagaaattca agcgttttat
69781 ctttagaaat cacagctagc atatgctgag aattactctc catggaaact tatactaaga
69841 ttgttttttt ccctcatatt taagccacta aagtcaaaag attagtacat tgacaactaa
69901 gtttagatgc tctatgcgga gaatcaattt catatgaatg tatcaagcaa ttcatgaact
69961 ctaggagacc ataaaatcca attgacagaa aaaatgagtc aactaacata tttacctgtg
70021 atatgaggta catgtgcagg tcaaagatca gaagaaaatt ttctccatga gtctcttgag
70081 cttccaactc atccagcgat ttgtatcaca aacaatctga aaaagaagct aaaaaacgtt
70141 ataccaaagt ttcacgccca taatgctatt gtttggttct ttcaagaacc tccccaatct
70201 tttgaattcg cattcaaaaa aaccatcagt gagtccattt caagtcggaa ctggcaggta
70261 ttattcatta tgacaaagta catacacttg ccccccactg aacaatgtca agaagggaaa
70321 acccgacatt gtgttggaat agctaaagtc tcatctcgtc tcgtgataca tgaaggttat
70381 caatatcaac ttgtagcaac tgtaatttac ttctaatatc tgataattct ttctggattc
70441 ctaaaagacg atcaagtctt agctgagctt cttctcgata aggcttggca acaatattca
70501 caaagttaac tagattactc gtcgcatctg aaagatcttt ttgcatagcg tcttcgagct
70561 gttgagccaa cgcatcagcc actttattca ccttaccaat tatagcctgt cttcgatatg
70621 ggaagtttgc tatagccaca tacctgtcac atagattatg ttatgcatac aaccagtctt
70681 tcttaaaagt cataaatatg cctctagttg caagaaaaaa atacactagg cgtgatctaa
70741 gaaggtggag taatgagaca ttgggaagag gggaaattta gagcagtgtt attaccctcc
70801 agcggagcaa aggccaagag caagaagatc ttccagtgtg gtcggtagca ctgaggttag
70861 aagtgatgca gacagtcctg cagctccaag cccaccaact gtcacaaaga actagaaaag
70921 gcaagcaaaa ctcaactatg tcaaaagtgt cacttagatt gattcttgaa tagcgagacg
70981 aagtatctgg gaaaatacgg tactgaatta acatctccgt cagatcatag gttcggattg
71041 aacagatgac acaattaaac aatgatgaag atcaagacac tttaatcgac tgaattc
```

Fig. 24

AtFzo-like Genomic Sequence

From
F15K9, AC005278:
F10O3, AC006550:

```
69001 aaaaactttt caaaacttca tgtgttgtga aaacaaaagt tttttggtaa tgaaaactcg
69061 acaaagacca gttaaaaacg tgtgtagtat aacttactgg taagtaaagc tataagcaag
69121 aatctgtacc ttatttttctc tctctctagt gagccctgac catccgaatt tcgcattcgc

69181 caatcgctgt gtttccgtgt gttttccccc tttttggttt tagatttgcc taaaccaatc
69241 agaacaagag aaacctggaa acaagaacca aaaaaagtgg gctttctctg catcatcatt
69301 ccacttctgg tccccaactg aaaaggacaa tccaaagcta gatcccttca aattttcctt
69361 tttgttttcg aaattttcgc aatttttaat attattttgg aagtctatgt ttctttctga
69421 tctttagcaa caaaggaagg tggaatctgt ttcacgttta cacaaaaaca tgtcaactgg
69481 agattttctc tttccctaac ttttgaccat acagtatggt ccatacttaa tattctctct
69541 ttgtttttaa taaaataaaa ggtttggtta tcaagcatat atgtcattag cttaaagcta
69601 tgactttgtt tagaaaactt aggaggacca tatggcaagc ttttatacag tgttagactt
69661 ctaacgttaa ttctaaacaa tctccagtat caagcattaa caaggtttat tctagcacct
69721 ctggattttt aaaacttctc gaaccaatcc ttaactaaaa aagaaattca agcgttttat
69781 ctttagaaat cacagctagc atatgctgag aattactctc catggaaact tatactaaga
69841 ttgtttttttt ccctcatatt taagccacta aagtcaaaag attagtacat tgacaactaa
69901 gtttagatgc tctatgcgga gaatcaattt catatgaatg tatcaagcaa ttcatgaact

69961 ctaggagacc ataaaatcca attgacagaa aaaatgagtc aactaacata tttacctgtg
70021 atatgaggta catgtgcagg tcaaagatca gaagaaaatt ttctccatga gtctcttgag
70081 cttccaactc atccagcgat ttgtatcaca aacaatctga aaaagaagct aaaaaacgtt
70141 ataccaaagt ttcacgccca taatgctatt gtttggttct ttcaagaacc tccccaatct
70201 tttgaattcg cattcaaaaa aaccatcagt gagtccattt caagtcggaa ctggcaggta
70261 ttattcatta tgacaaagta catacacttg ccccccactg aacaatgtca agaagggaaa
70321 acccgacatt gtgttggaat agctaaagtc tcatctcgtc tcgtgataca tgaaggttat

70381 caatatcaac ttgtagcaac tgtaatttac ttctaatatc tgataattct ttctggattc
70441 ctaaaagacg atcaagtctt agctgagctt cttctcgata aggcttggca acaatattca
70501 caaagttaac tagattactc gtcgcatctg aaagatcttt ttgcatagcg tcttcgagct
70561 gttgagccaa cgcatcagcc actttattca ccttaccaat tatagcctgt cttcgatatg
70621 ggaagtttgc tatagccaca tacctgtcac atagattatg ttatgcatac aaccagtctt
70681 tcttaaaagt cataaatatg cctctagttg caagaaaaaa atacactagg cgtgatctaa
70741 gaaggtggag taatgagaca ttgggaagag gggaaattta gagcagtgtt attaccctcc
70801 agcggagcaa aggccaagag caagaagatc ttccagtgtg gtcggtagca ctgaggttag
70861 aagtgatgca gacagtcctg cagctccaag cccaccaact gtcacaaaga actagaaaag
70921 gcaagcaaaa ctcaactatg tcaaaagtgt cacttagatt gattcttgaa tagcgagacg

70981 aagtatctgg gaaaatacgg tactgaatta acatctccgt cagatcatag gttcggattg

71041 aacagatgac acaattaaac aatgatgaag atcaagacac tttaatcgac tgaattc

                          tagttagaac ttagactaaa agtatttaat acttgaagct
  241 caccacttct cgaatatctt gttccaatcg ttttgatgtg gttccggcac tcaagttctg
  301 tattgttttc aagctgactt tatcagtttt ctgaagtaag tcatatgtgt ctatgcccaa
  361 ttgcgttttt gaattgacat atgttggcca tttgttttcg aatgatttca gagacagact

  421 cccttcacgg gcagtatttg attgtagcca ttcagcatat tttccaagca attcctgcaa
  481 acagtgaaat gtaaagtcaa tcaggtcaca acaagacatt gttagacaat atttactttc
  541 tgcatgaata gtgactatat ctcagacctc atatatatga ccaacatgtc cccagttagg
  601 ccaatactca aagaataaag catcacactt actttcgcat ttgtgagtgc tggagcgagt
  661 atttcacctt gaactttgga tgtagctgct actgaggccg atttttcccc tttgaacacg
  721 tacgagatcg caagatcaag gcttgatagt cgcagggtag ttcctatcag atcaacaact
  781 tgtaatctgg cattatcaat ctgcgataaa agccagagta caaaacacaa aaaagccaag

  841 ttagaacaat ccaatttcct ccttcgtgat tcaacaagat aatatctaat agaatttata
  901 ccaacgagag agcctgcctt ctccaagaaa tgctctcata ttccatctta agcgcgtatt
  961 ctttagttcg actgataatc ttgtctgctg aagccaagtc ttccctagca gctaggcaat
```

Fig. 24 continued 2/3

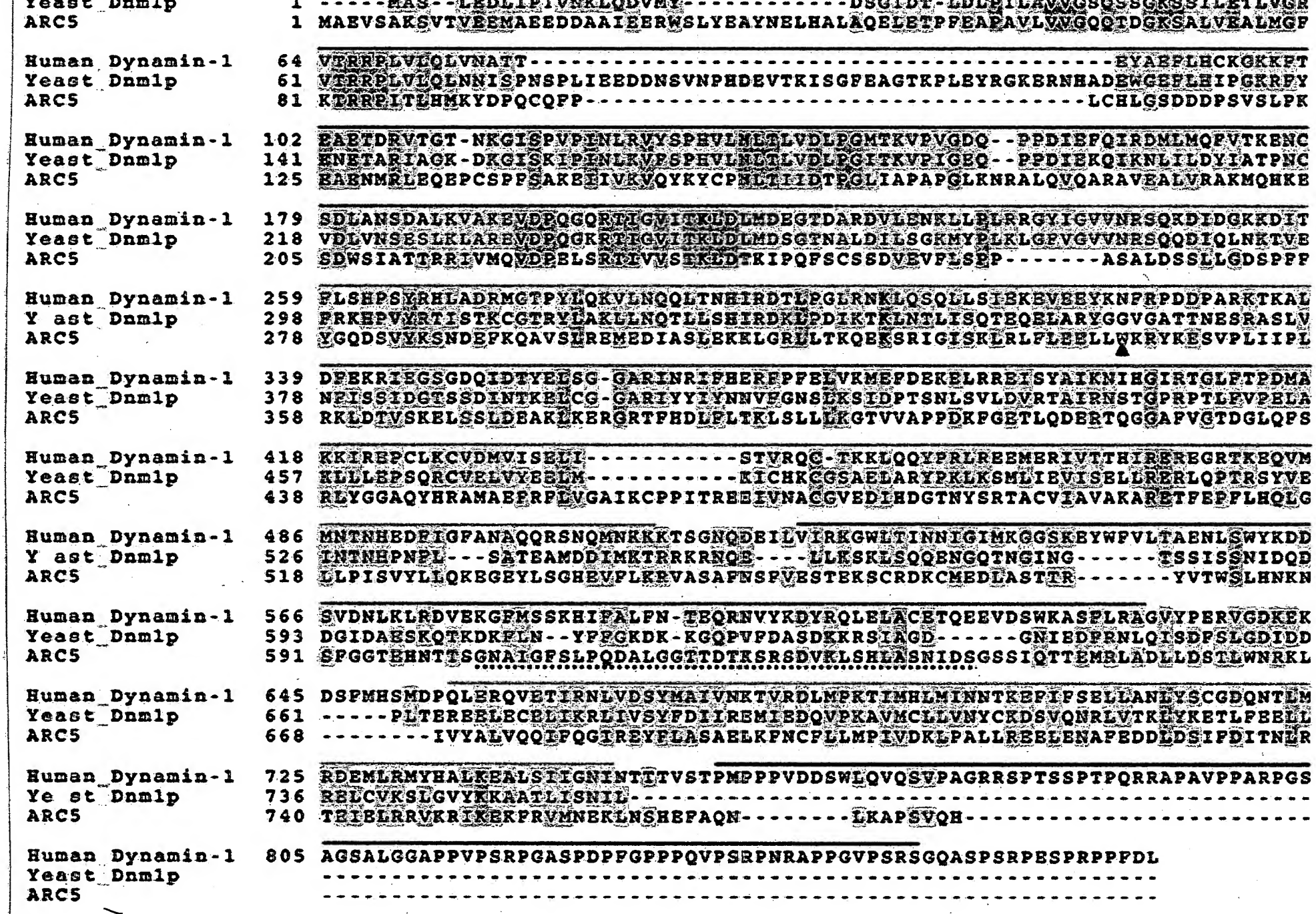
```
1021 cttgtctcac aagagcttcc acagaagaaa ggagacgctc agcaatcgcc atgggtgtct
1081 ccaatttaag ccttattctc tccatcccgg tagctgttga gctatccaag aagctataaa
1141 gaaatttctc aagttcattg aagctctgga ctctccaatt agaaccagga tctgcgatct
1201 caagatcatc tctgccaacc aaagaagctg ttgaaagctt cgcctcaaga gcagaccgtg
1261 cggacaccgg atacaatatc acattttctg tattaagcaa cttccgtgta ttctctttaa
1321 caaatgaaat agcttcctca agctgcagca gtaaaattaa tctttagtca agaaaaatcc
1381 agcaattaca gagactatca aggaaaagac ataaatattg tttctgataa acctcacgag
1441 catcacgata gatatcagat ttattcagaa taaacacaaa tttctttttc cactgctgtg
1501 tgtaccggag aaacgcaacc tgacaaacgc aatgagatgt aaatcagcct acagtaaaat
1561 caagacagca ggaagctcca gaggtagaga tagaaatgac atgggtatga tgacattgag
1621 aagcttattc cttcttttga aacttttca gattttcaaa ataagaacca aatattcata

1681 aggttagtac ttcgtagatg atcatttaca tacaatttgc catggacgta aaatcagttt
1741 tgctgtacca ataacaactc acctctcaaa tgtattggcc taagctttac caaaatatgg
1801 accaacagga catactggtt ccagcaatag ctataacact cattactttg aaaataaatt
     1861 agtctagttt aaatttataa tcaacaaaac cttgatatat gcattcaacc tagtgacaac
1921 tttcccttcc ctacacaagg tcttgatggt ctatacatga cttaactaac tactttcgtg
1981 aaaaatagat ctaacaagct acttaggcag atacatattg agcaaaaaca acaacaacta
2041 acatgccaag taaaacggta acttctacct cactttcagt taaagggcgg tcagcagaaa
2101 gaacaaaaac aagcaaatct gcacgtggaa caaattcttc tgtaagacgc tgttgccttt
2161 gaaggatcac attggtccca ggtgtgtcaa caatatttat ctgctcgaaa ggagaacaac
2221 ccaaatcaat tcctccaaat aattagtaaa aattcaggac tgcgatggta gaattttgaa
2281 ttactcacat ccttaagtat tggtgcagga agatagcata catattggcc atctggatgt
2341 gtttggcaac gttgttgctc ttcggattcc aagtcagagt agcacagaaa cgtgatttca
2401 ttggtagtgg ggactacccc ttctttcagg tatctcttcc caagaagtgc attgataacc
2461 gttgattttc cagagttaaa ttccccctag taccagcaaa ttgtggattc caattaatta
2521 tccatacaaa aactattttg aatcataaag gcaacttgtc ccacttaaaa catcttccaa
2581 gaacggcata ctcacaagtc acaagaaagc gaaaaaacta tgtacacaag cagatggtac
2641 agttaagaaa tatgcagttc ccttacaacc agatccataa gcaaattaag aagagaagca
2701 tagacagaac aaaagaagaa gcgaaacatc atagtcatag gcgacgaaag aagatattga
2761 gaagcactaa aggacaagtg actagaagca tgcccattgt atctcaagct atctagtgct
2821 ggcaaaagtg tgtaactaaa atgtttctgg aatgaatata gagctaatac caatgcaaat
2881 tcctggaatc acaatttgac ggagttgagt gcagaattac cactataacc atcagaaacg
2941 gctcatcgat ccgagaaaca gcatcaatta gaagggagac ttcctccatc tgttaatgaa
3001 aaatagtcga agaagtaaat gaaaccatat tataaccaca caaaaaccta cttagaacca
3061 tttcaaaaaa ttgacttcaa ctaccaaatg tttgaaataa aaatcaccag tggagccgcc
3121 ttgtggatga tttcaatcgt ctctctcaac acagatttct ccatttctac tattagtttc
3181 tgtttgtcct ctaattttat gaagccagca gaattctttt tctccggcag ttcattcata
3241 ttttgtgtct catgattatt tacaacataa gctccatcaa gactctggcg aagagctaca
3301 tccctagaag aacgcaaatc tttcaacgaa ataacaaaac cagaaacacc tgatttcagt
3361 aactgcaatt cttctttagc ttcttcattg cctctgcaag tcacatatat cggtattttc
3421 acgctcttca acaaagaatc cgccacttgc gtatcttctt caccagatcc aagtataagg
3481 aaatcagcac cctcggagct tgaggcaatt agagcagaat caacatcctt cacaatccga
3541 gctaccagtg gaagaagtac cgagtcgggg ttggatccca tcaatgtgtt tctcgccaca
3601 atcgccggaa gacctgtaca tagcattgag agacaaaact cactctatga gaaattaaat
3661 gctaaaactg aatgaaatca gttgttacct tcgtcggaga gagcaacacc actagcacca
3721 acggcggagg cgatatcaac acgttcagcg atcaagaggt aagcacggcc tttgacaagt
3781 gatttcagca aacaagccgc ctcgtagagc ttaccagcgg tggctccgcc atcaatcacg
3841 acgatttgaa ccgatttagc taaagcacgg tcgaccaaat caagagtctc ttcacgattc
3901 ccgctcataa cctcgtcggc gtctagccgg agaagtaaac cgggaacggc gagttcggga
3961 cgcttgtaac caccaggata aagagttcgc ggcctagaag aagaagtctg atcggcggat
4021 tcatgcgaaa tgtttctgat cgagagagaa gaaaaacgcc tatgacgtgg aggagtaaag
4081 gaggataact taaagcaccg gccaggaaac ggtggagatg cggcggagat aagaaacggt
4141 gacgtcacac attgccggtg agagattaga gttctcatgg agagagaaag agagagagtg
4201 agataccttc gtctgtgttt ctgattagtt ttcttatctc ttaaatatcc tattggccca
4261 tcccaaagct ataagacgga cggcgcgatt aattactttt caaacacatg aattaacgtt
4321 ttcacatatg tgttcatatc caaaaggtcc aaagtatacc acgaaaaggg agaaaaacag
4381 atttaaattc gtgaaatccc tctcccacaa ttaaatttac ttcttccaaa caaagacaaa
4441 cggcttgaac cagtcaagta agtgatacgg caccactaga tgttccagag cctccatctt
4501 ttttaatacg aagaagattt gtcctttgtg tgtatgaatt taacaagttt taattataga
4561 tttgtgtgtg tatgaattta aaaacctagt acgtagcatc agggaatgat atcatagcta
4621 ttttagttga gctttcaaat aagagatgat caaaatttag aacttctaag aacatgaacg
4681 aataaacaac tattttcttt tcaaaccaac taaggtagat ggtcactgaa agtatataca
4741 tcagataaaa gttgcttgtt attccagatg aagttggacc gagaaaaaaa aaagttactt
4801 gttattcaat atgtttggat ctttgtcttg cagattgcta tatagggttg ataatgggct
4861 tcgttgtaat gggtatacag tgtataagaa tcggccttgt gcaaccaatc ctaatatgtg
4921 tgtctcatta aggtaagtgc ttaagattag aagagtaaaa cacttgactt atcaactatg
4981 tcaactaagg gttctatatt tttattaaat aaaaaataat tgaatatttt ttagaatgat
5041 ttaataaatt taatgctatt gtttgattta aatgtataat tcaccgcgag aagaaatttt
```

149

Fig. 24 continued 3/3

```
5101 ataactcaaa ttttaaagtt ttaagttgta tttgtttatt ttgttaaatg tttaatattg
5161 tataattgta ttttgattgt tgtttctcgg atttcacccg tagtacatca tcccatatta
5221 atatcgaatc aaacccgtca attctaaaat ttcacccgtg gtagtattta attgtataat
5281 tatattttaa ttgtcattct aagatttcac tcctaattct atcgcaaatt attatcaacc
5341 caaaccagtc aattctaaaa tatcacccgt agtacaccat cccatattaa tatcgaatca
5401 agcccgtcaa ttctaggatt tcacccgtgg tagtatttaa ttgtataatt atattttaat
5461 tgtcattcta ggatttcact cctaattcta tcgcaaatta ttatcaaccc aaaccagtca
5521 attctaaaat atcacccgta gtacaccatc ccatattaat atcgattcaa actcgtcaat
5581 tctaggattt cgctcgtggt agtatttaat tgtataatta tattttaatt gtcattttaa
5641 ctcctagttc tatcgcaaat tcttatcaac ccaaacagtc aattctaaaa tttcacccgt
5701 agtataaagt ttaaatattt ataatattta aatttcttat aaaagaatca aaatgtgttt
5761 taaaaaaatt aaagttttaa gttttttttt tttaatattg ttaattttgt ttagtgttta
5821 agattatata attacattat gattgtcatt atatgttttt ctccatagca tactatccca
5881 tgttattatC CACTCAAACC TGTCACACCA TATAACcccg tcccgtgaaa ttaaacacaa
5941 atttgtcatt ttattataaa tttcaaatat ttataaaatt agaaacttca aaaaagatta
6001 atattgaccc aaacttcatc attgaatttt gagtgttata tctaagattt ctctcgcaat
6061 atatcgtccc gtattaatat cttttatatt gtttaaattt cttgtaaaat ttaatttata
6121 atttttaaa ctttttaaag tttcaatttt ttaaaataaa taaccctagg aaacaaacca
6181 ttttaattta aagataaact ttataaaaag tttttaaaat tataatattt aacttttgat
6241 aaagttataa tatttataat ttcttgaaac attttaaagt ttcaattctt taaaataata
6301 aatccgagta aaatcagata actattttaa ttttggacgc ttgataaatc aagcttcctg
6361 ctcattcgta atcagaatca ttttggtcct tttataatat gggtctgaac cattgtccaa
6421 tttttctaag cgatgtggga cattgtacac atattatttc ttcataggtt gaataatata
6481 tgtccgttta aaaaactttg aattacatca tattcagaaa aaaatataat attttattaa
6541 ctatatatat tttatataaa ttcaaaataa ataaagtata agatcaaata aaaatgaaag
```


Yeast Dnm1p 1 -----MAS--LEDLIPIVNKLQDVM----------DSGIDT-LDLPILAVVGSQSSGKSSILETLVGR
ARC5 1 MAEVSAKSVTVEEMAEEDDAAIEERWSLYEAYNELHALAQELETPFEAPAVLVVGQQTDGKSALVEALMGF
Human Dynamin-1 64 VTRRPLVLQLVNATT----------EYAEFLHCKGKKFT
Yeast Dnm1p 61 VTRRPLVLQLNNISPNSPLIEEDDNSVNPHDEVTKISGFEAGTKPLEYRGKERNHADEWGEFLHIPGKRFY
ARC5 81 KTRRPITLHMKYDPQCQFP----------LCHLGSDDDPSVSLPK
Human Dynamin-1 102 EAEIDRVTGT-NKGISPVPINLRVYSPHVLNLTLVDLPGMTKVPVGDQ--PPDIEFQIRDMLMQFVTKENC
Yeast Dnm1p 141 ENEIARIAGK-DKGISKIPINLKVFSPHVLNLTLVDLPGITKVPIGEQ--PPDIEKQIKNLILDYIATPNC
ARC5 125 EAENMRLEQEPCSPFSAKEIIVKVQYKYCPNLTIIDTPGLIAPAPGLKNRALQVQARAVEALVRAKMQHKE
Human Dynamin-1 179 SDLANSDALKVAKEVDPQGQRTIGVITKLDLMDEGTDARDVLENKLLPLRRGYIGVVNRSQKDIDGKKDIT
Yeast Dnm1p 218 VDLVNSESLKLAREVDPQGKRTIGVITKLDLMDSGTNALDILSGKMYPLKLGFVGVVNRSQQDIQLNKTVE
ARC5 205 SDWSIATTRRIVMQVDPELSRTIVVSEKLDTKIPQFSCSSDVEVFLSEP-------ASALDSSLLGDSPFF
Human Dynamin-1 259 FLSHPSYRHLADRMGTPYLQKVLNQQLTNHIRDTLPGLRNKLQSQLLSIEKEVEEYKNFRPDDPARKTKAL
Y ast Dnm1p 298 FRKHPVYRTISTKCGTRYLAKLLNQTLLSHIRDKLPDIKTKLNTLISQTEQELARYGGVGATTNESRASLV
ARC5 278 YGQDSVYKSNDEFKQAVSLREMEDIASLEKELGRLLTKQEKSRIGISKLRLFLEELLVKRYKESVPLIIPL
Human Dynamin-1 339 DFEKRIEGSGDQIDTYELSG-GARINRIFHERFPFELVKMEFDEKELRREISYAIKNIHGIRTGLFTPDMA
Yeast Dnm1p 378 NFISSIDGTSSDINTKELCG-GARIYYIYNNVFGNSLKSIDPTSNLSVLDVRTAIRNSTGFRPTLFVPELA
ARC5 358 RKLDTVSKELSSLDEAKLKERGRTFHDLFLTKLSLLLKGTVVAPPDKFGETLQDERTQGGAPVGTDGLQPS
Human Dynamin-1 418 KKIREPCLKCVDMVISELI----------STVRQC-TKKLQQYPRLREEMERIVTTHIRERGRTKEQVM
Yeast Dnm1p 457 KLLLEPSQRCVELVYEELM----------KICHKCGSAELARYPKLKSMLIEVISELLRERLQPTRSYVE
ARC5 438 RLYGGAQYHRAMAEFRFLVGAIKCPPITREEIVNACGVEDIHDGTNYSRTACVIAVAKARETFEPFLHQLG
Human Dynamin-1 486 MNTNHEDFIGFANAQQRSNQMNKKKTSGNQDEILVIRKGWLTINNIGIMKGGSKEYWFVLTAENLSWYKDD
Y ast Dnm1p 526 INTNHPNPL---SATEAMDDIMKTRRKRNQE----LLKSKLSQQENGQTNGING------TSSISSNIDQD
ARC5 518 LLPISVYLLQKEGEYLSGHEVPLKRVASAFNSFVESTEKSCRDKCMEDLASTR-------YVTWSLHNKN
Human Dynamin-1 566 SVDNLKLRDVEKGFMSSKHIFALFN-TEQRNVYKDYRQLELACETQEEVDSWKASFLRAGVYPERVGDKEK
Yeast Dnm1p 593 DGIDAESKQTKDKFLN--YFFGKDK-KGQPVFDASDKKRSIAGD------GNIEDFRNLQISDFSLGDIDD
ARC5 591 SFGGTEHNTTSGNAIGFSLPQDALGGTTDTKSRSDVKLSHLASNIDSGSSIQTTEMRLADLLDSTLWNRKL
Human Dynamin-1 645 DSFMHSMDPQLERQVETIRNLVDSYMAIVNKTVRDLMPKTIMHLMINNTKEFIFSELLANLYSCGDQNTLM
Yeast Dnm1p 661 -----PLTERFELECELIKRLIVSYFDIIREMIEDQVPKAVMCLLVNYCKDSVQNRLVTKLYRETLFEELL
ARC5 668 ---------IVYALVQQIFQGIREYFLASAELKFNCFLLMPIVDKLPALLREELENAFEDDLDSIFDITNLR
Human Dynamin-1 725 RDEMLRMYHALKEALSIIGNINTITVSTPMPPPVDDSWLQVQSVPAGRRSPTSSPTPQRRAPAVPPARPGS
Ye st Dnm1p 736 RELCVKSLGVYKKAATLISNIL----------
ARC5 740 TEIELRRVKRIKERFRVMNEKLNSHEFAQN--------LKAPSVQH----------
Human Dynamin-1 805 AGSALGGAPPVPSRPGASPDPFGPPPQVPSRPNRAPPGVPSRSGQASPSRPESPRPPPDL
Yeast Dnm1p ----------
ARC5 ----------

191

Fig. 25

FIG. 26

ARC5 Homologous Sequences

<A class=dblinks
href="javascript:PopUpMenu2_Set(Menu22246438,",",",",");"

target=_self>Links</A></SPAN></TD></TR></TBODY></TABLE></DT></DL><PRE>L
OCUS BQ860973
712 bp mRNA linear EST 14-AUG-2002
DEFINITION QGC17C24.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
QGC17C24, mRNA sequence.
ACCESSION BQ860973
VERSION BQ860973.1 GI:22246438
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM <A href="http://www.ncbi.nlm.nih.gov/htbin-
post/Taxonomy/wgetorg?name=Lactuca+sativa
">Lactuca sativa</A>
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 712)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
<A
href="http://compgenomics.ucdavis.edu/">http://compgenomics.ucdavis.edu/</A>/
JOURNAL Unpublished
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: <A href="mailto:akozik@atgc.org">akozik@atgc.org</A> <A href="
mailto:[michelmore@vegmail.ucdavis.edu">[michelmore@vegmail.ucdavis.edu</A>]
singleton, see <A href="http://cgpdb.ucdavis.edu/">http://cgpdb.ucdavis.edu/</A>/
for
details.
Plate: QGC17 row: C column: 24.
FEATURES Location/Qualifiers

```
   source          1..712
                   /organism="Lactuca sativa"
                   /mol_type="mRNA"
                   /cultivar="Salinas"
                   /db_xref="<A href="
http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg?id=4236">taxon:4236</A>"
                   /clone="QGC17C24"
                   /lab_host="E.coli"
                   /clone_lib="QG_ABCDI lettuce salinas"
                   /note="Vector: pBRcDNASfiAB; The library was constructed
                   from 10 different sources of RNA from a single genotype.
                   Separate cDNAs were generated using primers that
                   incorporated unique 5' and 3' tags to distinguish each
                   source of RNA. cDNAs were then pooled, size-fractionated,
                   directionally cloned into a custom medium-copy vector and
                   transformations made with four size classes to minimize
                   size bias. Details of each source of RNA and library
                   construction can be obtained at <A href="http://cgpdb.ucdavis.edu/">
http://cgpdb.ucdavis.edu/</A>/
                   TAG_LIB=QG_ABCDI lettuce salinas
                   TAG_TISSUE=chemical induction
                   TAG_SEQ=TGTAGCCGGG"
BASE COUNT     206 a    152 c    142 g    210 t      2 others
ORIGIN
        1 ttgttcagct ccgccaaaag aatccaagaa ttggcgtaat ccggctcgat tcttattgtg
       61 aagggaccag gtgacataac gggtggtgct tattagatct tccatgcatt tttcatggca
      121 tgatctttcg gtggattcag caaagttata gaaagcagat gaaacacgtc tcaagaaaac
      181 ttcatggcca cttaggaatt cgccttcttt ctgaagaaga taaacggaga tgggaagtaa
      241 tctcttgaga atgtgaagaa gtcgactgcc caactgatga agaaaaggtt caaaagtatc
      301 acgagctttt gcaacagcga tgacacatgc agtcctggag taatttgttc catcatgaat
      361 atcttcgacc ccacatgcat tcacaatttc ttcacgtgta attgcagggc attttatccc
      421 tccaacaaca aacctaaatt cagccatggc acgatgatat tgtgcacctc catatagacg
      481 catacctgca ttaggtatta gtttgtgtgg gaactgagag ccatcaatac cgattaatgc
      541 ccctccatta accctctcat cttgtagtgt ttccccaaat ttatctggag gtgcaacaac
      601 tgtccctntt catagcagtg ataacttggt aaggaaaaga tcatgaaaag atctcncttt
      661 ctcctttagt ttgacttcat ctaaagtgct gagttcttga tttatgtcat tt
//
</PRE>
<DL>
 <DT>
 <TABLE cellSpacing=0 cellPadding=0 width="100%">
  <TBODY>
  <TR>
   <TD><INPUT type=checkbox value=13371119 name=uid><B>2: </B>BG452325.
    NF086D06LF1F1047 ...[gi:13371119] </TD>
   <TD align=right><SPAN>
```

<SCRIPT language=JavaScript1.2>
<!--
var Menu13371119 = [

["Taxonomy","window.top.location='/entrez/query.fcgi?db=nucleotide&cmd=Display&dopt=
nucleotide_taxonomy&from_uid=13371119'","",""],

["Help","window.open('/entrez/query/static/popup.html','Links_Help','resizable=no,scrollbars
=yes,toolbar=no,location=no,directories=no,status=no,menubar=no,copyhistory=no,width=40
0,height=500');","",""]
]
//-->
</SCRIPT>
<A class=dblinks
href="javascript:PopUpMenu2_Set(Menu13371119,",",",",",");"

target=_self>Links</A></SPAN></TD></TR></TBODY></TABLE></DT></DL><PRE>L
OCUS BG452325
666 bp mRNA linear EST 16-MAR-2001
DEFINITION NF086D06LF1F1047 Developing leaf Medicago truncatula cDNA clone
NF086D06LF 5', mRNA sequence.
ACCESSION BG452325
VERSION BG452325.1 GI:13371119
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM <A href="
http://www.ncbi.nlm.nih.gov/htbin-
post/Taxonomy/wgetorg?name=Medicago+truncatula">Medicago
truncatula</A>
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 666)
AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula leaf library
JOURNAL Unpublished
COMMENT Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: <A href="mailto:gdmay@noble.org">gdmay@noble.org</A>

```
        Insert Length: 666   Std Error: 0.00
        Plate: 086  row: D  column: 06
        Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES            Location/Qualifiers
   source        1..666
                 /organism="Medicago truncatula"
                 /mol_type="mRNA"
                 /db_xref="<A href="
http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg?id=3880">taxon:3880</A>"
                 /clone="NF086D06LF"
                 /tissue_type="leaf"
                 /dev_stage="Pooled developmental"
                 /clone_lib="Developing leaf"
                 /note="Vector: Lambda Zap; Contains a mixture of very
                 young, developing, mature and senescing leaves."
BASE COUNT     201 a    163 c    147 g    154 t      1 others
ORIGIN
      1 atctaaagta acaaccacca caaaacacaa caatggagga agaaagagaa caccaccaac
     61 tcaaagacaa agaagaaaac gagtggcgtc tctacgaagc ttacaatgaa cttcacgcgc
    121 ttgctcaaga acttcacacg cctttcgacg cgccggcggt actggttgtg ggccaccaaa
    181 cagacgggaa gagcgcctta gttgaggctc taatgggctt ccagttcaac cacgtcggtg
    241 gtggcaccaa aacccgccgg cccattactc ttcacatgaa atatggccca cattgcgagt
    301 ctccttcttg ctatcttctt tctgatgatg acccttctct ttctcaccat atgtcacttt
    361 cccaaatcca gggttatatt gaagctgaga atgcgaggtt ggagcgtgac tcatgttgtc
    421 aattttcagc taaggaaata atcataaaag tggaatacaa atactgtccc aatctcacca
    481 taatagacac accaggatta gttgctcctg caccaggtcg taaaaatagg gcgatacagg
    541 cacaggcacg agcggtagag tcactcgttc gtgcaaaaat gcagcacaag gagttcatta
    601 tactctgtct tgaagattgt agtgattgga gcaatgcgac tacgangcgc gttgtaatgc
    661 aaattg
//
</PRE>
<DL>
 <DT>
 <TABLE cellSpacing=0 cellPadding=0 width="100%">
  <TBODY>
  <TR>
   <TD><INPUT type=checkbox value=14878353 name=uid><B>3: </B>BI270606.
    NF056G04FL1F1036 ...[gi:14878353] </TD>
   <TD align=right><SPAN>
    <SCRIPT language=JavaScript1.2>
<!--
var Menu14878353 = [

["Taxonomy","window.top.location='/entrez/query.fcgi?db=nucleotide&cmd=Display&dopt=
nucleotide_taxonomy&from_uid=14878353'","",""],
```

["Help","window.open('/entrez/query/static/popup.html','Links_Help','resizable=no,scrollbars
=yes,toolbar=no,location=no,directories=no,status=no,menubar=no,copyhistory=no,width=40
0,height=500');","","",""]
]
//-->
</SCRIPT>
<A class=dblinks
href="javascript:PopUpMenu2_Set(Menu14878353,",",",",",");"

target=_self>Links</A></SPAN></TD></TR></TBODY></TABLE></DT></DL><PRE>L
OCUS BI270606
663 bp mRNA linear EST 18-JUL-2001
DEFINITION NF056G04FL1F1036 Developing flower Medicago truncatula cDNA clone
NF056G04FL 5', mRNA sequence.
ACCESSION BI270606
VERSION BI270606.1 GI:14878353
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM <A href="
http://www.ncbi.nlm.nih.gov/htbin-
post/Taxonomy/wgetorg?name=Medicago+truncatula">Medicago
truncatula</A>
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 663)
AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library
JOURNAL Unpublished
COMMENT Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: <A href="mailto:gdmay@noble.org">gdmay@noble.org</A>
Insert Length: 663 Std Error: 0.00
Plate: 056 row: G column: 04
Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES Location/Qualifiers
source 1..663
/organism="Medicago truncatula"
/mol_type="mRNA"

```
                /db_xref="<A href="
http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg?id=3880">taxon:3880</A>"
                /clone="NF056G04FL"
                /tissue_type="Developing flowers"
                /dev_stage="Developmentally pooled. Contains a mixture of
                very young, developing, fully-opened flowers and flowers
                in early transition into pods."
                /clone_lib="Developing flower"
                /note="Vector: Lambda Zap; cDNA was prepared from polyA+
                enriched, pooled samples of equivalent amounts of total
                RNA from very young, developing, fully-opened flowers and
                flowers transitioning into pods. The cDNA was
                directionally ligated into the Uni-Zap XR vector
                (Stratagene) and packaged using the Gigapack III Gold
                packaging extracts. Phagemids containing cDNA inserts were
                in vivo excised from the recombinant Uni-ZAP XR vector
                using ExAssist helper phage and the E. coli strain
                XL1-Blue MRF' (Stratagene). Excised plasmids were plated
                using SOLR cells."
BASE COUNT     191 a    141 c    144 g    187 t
ORIGIN
      1 gtctttatgg gggtgcacaa tatcatcgag caatggctga atttcgtttt gtagttggag
     61 gaatcaagtg ccctccaatt acccgggaag aaattgtaaa tgcttgtgga gttgaagaca
    121 ttcatgatgg aacaaactac tctaggactg cttgtgtaat tgctgttgca aaggctcatg
    181 atacatttga acctttttctt catcagttgg ggtctagatt gttgcacata cttaagagat
    241 tgctcccaat ctcttttat cttcttcaga aagattgtga gtatctaagt ggccatcagg
    301 tgttcctcag gcgtgttgcc tccgccttcg acaactttgc agaatccact gaaaaatcat
    361 gccgtgaaaa atgtatggag gacttggtaa gcaccacacg atatgtctca tggtctctac
    421 acaataagag tcgggcagga ttacgccagt tcttagattc atttggtgga acagaacatt
    481 ccaatgtttg taatgatccc actgcaactg ttctatcaca aacaaatgtg caagagaagg
    541 aagacacaaa gccacaacta gaagtaaagc tcagtcacgt ggcctctgga actgatccta
    601 gcacatccac ccagacagct gaaacaaagc ttgctgacct tcttgatagt acactttgga
    661 atc
//
</PRE>
<DL>
 <DT>
 <TABLE cellSpacing=0 cellPadding=0 width="100%">
  <TBODY>
  <TR>
   <TD><INPUT type=checkbox value=22485477 name=uid><B>4: </B>BU045400.
    PP_LEa0022H05f Pe...[gi:22485477] </TD>
   <TD align=right><SPAN>
    <SCRIPT language=JavaScript1.2>
<!--
var Menu22485477 = [
```

["Taxonomy","window.top.location='/entrez/query.fcgi?db=nucleotide&cmd=Display&dopt=
nucleotide_taxonomy&from_uid=22485477'","","",""],

["Help","window.open('/entrez/query/static/popup.html','Links_Help','resizable=no,scrollbars
=yes,toolbar=no,location=no,directories=no,status=no,menubar=no,copyhistory=no,width=40
0,height=500');","","",""]
]
//-->
</SCRIPT>
<A class=dblinks
href="javascript:PopUpMenu2_Set(Menu22485477,",",",",",");"

target=_self>Links</A></SPAN></TD></TR></TBODY></TABLE></DT></DL><PRE>L
OCUS BU045400
622 bp mRNA linear EST 26-AUG-2002
DEFINITION PP_LEa0022H05f Peach developing fruit mesocarp Prunus persica cDNA
clone PP_LEa0022H05f, mRNA sequence.
ACCESSION BU045400
VERSION BU045400.1 GI:22485477
KEYWORDS EST.
SOURCE Prunus persica (peach)
ORGANISM <A href="http://www.ncbi.nlm.nih.gov/htbin-
post/Taxonomy/wgetorg?name=Prunus+persica
">Prunus persica</A>
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE 1 (bases 1 to 622)
AUTHORS Callahan,A., Palmer,M., Main,D., Wing,R. and Abbott,A.
TITLE Peach Model Genome for Rosaceae
JOURNAL Unpublished
COMMENT Contact: Abbott, A.
Dept of Genetics and Biochemistry
Clemson University
122 Long Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 3060
Fax: 864 656 6879
Email: <A href="mailto:aalbert@clemson.edu">aalbert@clemson.edu</A>
Total High Quality bases = 553
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 622.
FEATURES Location/Qualifiers
source 1..622
/organism="Prunus persica"
/mol_type="mRNA"

```
                /cultivar="Loring"
                /db_xref="<A href="
http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg?id=3760">taxon:3760</A>"
                /clone="PP_LEa0022H05f"
                /tissue_type="Mesocarp"
                /lab_host="E. coli"
                /clone_lib="Peach developing fruit mesocarp"
                /note="Vector: pBluescript II SK(-); Site_1: EcoRI;
                Site_2: XhoI; authority=Prunus persica L. Batsh; The
                sequence has been trimmed to remove vector sequence and
                contains a minimum of 100 bases of phred value 20 or
                above. For more details on library preparation and
                sequence analysis go to
                <A href="http://www.genome.clemson.edu/projects/peach">
http://www.genome.clemson.edu/projects/peach</A>. To order
                this clone go to <A href="http://www.genome.clemson.edu/orders">
http://www.genome.clemson.edu/orders</A>"
BASE COUNT    168 a    125 c    147 g    181 t      1 others
ORIGIN
      1 gcttatacct aacgcaggaa tgcgtttata tggtggtgca caataccacc gtgccatggc
     61 tgagttccgc tttgtagttg gaggaataaa atgccctcca attacaaggg aagaaattgt
    121 aaatgcatgt ggagttgaag atttacatga tggcacaaac tactcaagga cagcttgtgt
    181 aatagccgtt gcaaaggccc gtgatacatt tgagcctttc cttcatcagt taggttgtag
    241 actcttgcac attctaaaga gattacttcc tatatcagtc tatcttcttc agaaagatgg
    301 tgagtattta agtggccatg aggtgtttct taggcgtgtt gcttctgctt tcaatgactt
    361 tgcagaatct accgaaaggg catgtcgtga aaaatgcatg gaggatttag taagcaccac
    421 ccgctatgtc acctggtccc ttcacaacaa gaatcgagct gggttacgtc aatttttaga
    481 ctcgttcgct ggaacagaac ataacactat gggtagtaat tgcgtacctg ctggtatttc
    541 ccaagattca tcctttgggt ctgttgccaa tgagaaggat actaagtcaa gggcagatgt
    601 gaagctcanc catgtggcgt ct
//
</PRE>
<TABLE cellSpacing=0 cellPadding=0 width="100%" bgColor=#cccccc>
 <TBODY>
 <TR>
  <TD>
   <TABLE cellSpacing=0 cellPadding=0>
    <TBODY>
    <TR>
     <TD noWrap><INPUT onclick="GoV (')" type=button value=Display name="">
       <SMALL><SELECT onchange=form.view.selectedIndex=selectedIndex
      name=view1><OPTION value=DocSum>Summary</OPTION> <OPTION
       value=asn>ASN.1</OPTION> <OPTION value=est>EST</OPTION> <OPTION
       value=fasta>FASTA</OPTION> <OPTION value=fasta_xml>TinySeq
       XML</OPTION> <OPTION value=gb selected>GenBank</OPTION> <OPTION
       value=gb_xml>GBSeq XML</OPTION> <OPTION value=gi>GI List</OPTION>
```

```
          <OPTION value=graph>Graphics</OPTION> <OPTION
          value=xml>XML</OPTION> <OPTION
         value=def>default</OPTION></SELECT></SMALL><SMALL>
           Show: </SMALL><SELECT
         onchange=form.dispmax.selectedIndex=selectedIndex
          name=dispmax1><OPTION value=1>1</OPTION> <OPTION
          value=2>2</OPTION> <OPTION value=5>5</OPTION> <OPTION
          value=10>10</OPTION> <OPTION value=20 selected>20</OPTION> <OPTION
          value=50>50</OPTION> <OPTION value=100>100</OPTION> <OPTION
          value=200>200</OPTION> <OPTION value=500>500</OPTION></SELECT>
           <INPUT onclick="GoV
(form.SendTo.options[form.SendTo.selectedIndex].value,4)" type=button value="Send to"
name="">
          <SELECT onchange=form.SendTo.selectedIndex=selectedIndex
         name=SendTo1><OPTION value=on selected>File</OPTION> <OPTION
          value=t>Text</OPTION> <OPTION
          value="Add to Clipboard">Clipboard</OPTION></SELECT>
    </TD></TR></TBODY></TABLE></TD></TR>
 <TR>
  <TD>
   <TABLE cellSpacing=0 cellPadding=0 width="100%">
    <TBODY>
    <TR>
     <TD align=middle width="50%">
      <DIV class=medium2>Items 1-4 of 4</DIV></TD>
     <TD align=right width="100%">One page.</TD></TR><INPUT type=hidden
     value=20 name=showndispmax><INPUT type=hidden value=0
    name=page></TBODY></TABLE></TD></TR></TBODY></TABLE></FORM><BR>
<DIV class=medium1 align=center>
<P><A href="http://www.ncbi.nlm.nih.gov/About/disclaimer.html">Disclaimer</A> |
<A href="mailto:info@ncbi.nlm.nih.gov">Write to the Help Desk</A><BR><A
href="http://www.ncbi.nlm.nih.gov/">NCBI</A> | <A
href="http://www.nlm.nih.gov/">NLM</A> | <A href="http://www.nih.gov/">NIH</A>
</P>
<P> </P></DIV>
<P class=dblinks align=right><FONT color=#eeeeee size=-5>Jun 19 2003 12:37:45 <!--
ipubmed7
--></FONT></P>
<SCRIPT language=JAVASCRIPT> /* <!-- */ TextFocus (); // -->
</BODY></HTML>
```

FIG. 27

Fzo-like Homologous Sequences

1: BG890612. EST516463 cSTD So...[gi:14267734]

LOCUS BG890612 752 bp mRNA linear EST 07-MAR-2003
DEFINITION EST516463 cSTD Solanum tuberosum cDNA clone cSTD19A23 5' sequence, mRNA sequence.
ACCESSION BG890612
VERSION BG890612.1 GI:14267734
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 752)
AUTHORS van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE Generations of ESTs from dormant potato tubers
JOURNAL Unpublished
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
FEATURES Location/Qualifiers
source 1..752
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD19A23"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/clone_lib="cSTD"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored

for one month post-harvest at 4oC. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."

BASE COUNT 226 a 144 c 172 g 210 t
ORIGIN

```
   1 gcgaatgtga ttcttcaaag gcaacaaagg ctgacggagg aatttgtgcc tcgtgcagat
  61 ctgcttctgt ttctcatgtc tgctgatcga ccattaactg aaagtgaggt tagttttctg
 121 cgttacactc agcagtggag taagaaggtc atttttgtgc tgaacaagtc tgacatatac
 181 aagaataacg gcgagttgga ggaggccatt gcatttatca aagaaaatac acggaaattg
 241 ctgaatacag aatccgtaac actgtatcca gtatctgcac ggctcgctct tgaatcaaag
 301 ctttctactt ttgatggtgc ccttagtcaa aacaatggga gttcaaataa tgattctcac
 361 tggaaaacca agagcttcta tgagcttgag aagtacttgt ctagcttttt ggattcatcc
 421 acaagtactg gaattgagag aatgaagctg aagcttgaaa ctccaattgc cattgcagaa
 481 caactacttt tagcttgtca aggacttgtg agacaagaat gtcagcaagc caaacaagac
 541 ttgctgtttg ttgaggatct tgtcaacagc gtagaagagt gcacaaagaa gctggaagtt
 601 gatagcattc tgtggaagag gcaggttcta tctctgataa actctgctca agcacgtgtt
 661 gtccggcttg tagagtcaac gttacaactg tcaaatgttg atcttgtcgc tacatatgta
 721 ttcagaagag aaaactctac tcaaatgcca gc
//
```

2: AW760673. sl53d10.y1 Gm-c10...[gi:7692570]

Links

LOCUS AW760673 492 bp mRNA linear EST 03-DEC-2001
DEFINITION sl53d10.y1 Gm-c1027 Glycine max cDNA clone GENOME SYSTEMS
CLONE ID:
Gm-c1027-5036 5' similar to SW:YOR6_CALSR P40983 HYPOTHETICAL PROTEIN IN XYNA 3'REGION ;, mRNA sequence.
ACCESSION AW760673
VERSION AW760673.1 GI:7692570
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 492)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.

TITLE Public Soybean EST Project
JOURNAL Unpublished
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 2209 Std Error: 0.00
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1..492
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1027-5036"
/tissue_type="cotyledons of 3- and 7-day-old Williams seedlings"
/lab_host="DH10B"
/clone_lib="Gm-c1027"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 3- and 7-day-old Williams seedlings which were propagated on paper towels with distilled water. The cotyledons were flash-frozen in liquid nitrogen, then lyophilized for 72 hours. Unequal amounts of mRNA was used for cDNA synthesis. Stratagene's cDNA Synthexix Kit (catalog number 200401) was used to synthesize the cDNA. First- stranded synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An anchor nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18] to anchor the primer at the 5' end of the poly(A) tract. After second- strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff,

using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript(tm) II XR Predigested vector (pBluescript II SK(+) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 97% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=30). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 135 a 91 c 108 g 158 t
ORIGIN
1 tgttgaatga agctattgaa gctatcaaga gggctgcacc tctgatggag gaggtttcac
61 ttcttaatga tgcggtttct caaattgatg agccattctt actggttata gtgggggaat
121 tcaactctgg taaatctacc gtgattaatg cgcttcttgg agaaagatat ctcaaagagg
181 gagttgttcc aacaactaat gagatcacat ttttacgata tactgactta gatattgaac
241 aacaacggtg tgaaaggcat ccagatggcc aatatatttg ctacattcct gctccaattc
301 ttaaagagat gaccattgtt gatacacctg gaactaatgt gattcttcag aggcagcagc
361 gtcttacaga ggaatttgta ccccgtgcag atttacttct ttttgtcatt tctgctgatc
421 gcccttttaac tggaagtgag attgcttttc ttcgttattc tcagcagtgg aaaaagaaag
481 cggtctttgt ct
//

3: BE353824. EST355167 tomato ...[gi:9291800]

Links

LOCUS BE353824 446 bp mRNA linear EST 18-MAY-2001
DEFINITION EST355167 tomato flower buds, anthesis, Cornell University Lycopersicon esculentum cDNA clone cTOD6M4, mRNA sequence.
ACCESSION BE353824
VERSION BE353824.1 GI:9291800
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 446)
AUTHORS van der Hoeven,R.S., Bezzeredes,J.L., Matern,A.L., Holt,I.E., Liang ,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman ,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
TITLE Generation of ESTs from tomato flower tissue, anthesis
JOURNAL Unpublished
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University

100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES Location/Qualifiers
source 1..446
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOD6M4"
/tissue_type="flower"
/dev_stage="anthesis"
/clone_lib="tomato flower buds, anthesis, Cornell University"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."
BASE COUNT 119 a 82 c 116 g 129 t
ORIGIN
1 gagaccatta agtacaattc tataagcagt cttttgaaaa aagatggact tcattggtga
61 atccgtctga ccaaattgag ttaggaacaa ctggtgtgct ggatagaaaa tctgaagtta
121 ccataagtgt catagaggat ttcagtgctg cagctgcttc aaaattgctt gagagagata
181 ttcgtgaagt gttcttgggt acttttggtg gtcttggagc agctggttta tcagcgtcgc
241 ttctgacatc tgttcttcaa accacattag aagacctcct tgcacttggc ctttgttctg
301 ctggcgggtt attagcggtc ttcaacttct catcccggag acagcaagtg gtagataaag
361 taaagaggac tgctgatggc ctttcacgtg aactcgaaga ggctatgcag aaggagctct
421 tggagacgac tagtaatgtg gaggac
//

4: BI136291. F066P17Y Populus ...[gi:18017219]
Links

LOCUS BI136291 521 bp mRNA linear EST 31-DEC-2001
DEFINITION F066P17Y Populus flower cDNA library Populus balsamifera subsp. trichocarpa cDNA, mRNA sequence.
ACCESSION BI136291
VERSION BI136291.1 GI:18017219
KEYWORDS EST.
SOURCE Populus balsamifera subsp. trichocarpa
ORGANISM Populus balsamifera subsp. trichocarpa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Malpighiales; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 521)
AUTHORS Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hiltonen ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bahlerao,R., Jansson,S., Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and Lundeberg,J.
TITLE Gene expression in Populus
JOURNAL Unpublished
COMMENT Contact: Erlandsson R
Department of Biotechnology
Royal Institute of Technology
Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
Email: rikerl@biochem.kth.se.
FEATURES Location/Qualifiers
source 1..521
/organism="Populus balsamifera subsp. trichocarpa"
/mol_type="mRNA"
/sub_species="trichocarpa"
/db_xref="taxon:3694"
/clone_lib="Populus flower cDNA library"
/note="Organ: flower"
BASE COUNT 143 a 87 c 135 g 156 t
ORIGIN
1 tggtgttgtg ctgtctgatc aagggcttcc tgcccttgtg gcaagaaata tgatgatggg
61 ttctcgaact gaatcagttg ttctaccttt ggtagccagg attgtgcaga caccatatgc
121 tgcattaaat gcgtctaatt ctgaaggtgc tgattttctt atatatgttc atggcccaga
181 ggatgatcct gatgtagaaa tgagccctgg attcgggaat gtgaagatac caatctttgt
241 cctcaatgct tcacgtgggg aggacacatt gtcggtgggg gcatcaaaat ttctgaaaac
301 cggtgctagt ggtttagttc tgtcattgga agatttgagg ttatttagcg atgatgcttt
361 gagtcagatg tttgacactc tgagtgcaac cggtaaaaac tttcaggatg accttgaaag
421 cttcagtaag ctcaaatcta tggatatgga aaatgatatt catgaaaaaa caacggtggc
481 aggctttgtt aaactggagg atagagaaaa acagctcata g